American Economic Review 2009, 99:4, 1384–1414
http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.4.1384

The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial

By Joshua Angrist and Victor Lavy*

The Israeli matriculation certificate is a prerequisite for most postsecondary schooling. In a randomized trial, we attempted to increase certification rates among low-achievers with cash incentives. The experiment used a school-based randomization design offering awards to all who passed their exams in treated schools. This led to a substantial increase in certification rates for girls but had no effect on boys. Affected girls had a relatively high ex ante chance of certification. The increase in girls' matriculation rates translated into an increased likelihood of college attendance. Female matriculation rates increased partly because treated girls devoted extra time to exam preparation. (*JEL* I21, I28, J16)

> *Thus, the teacher may say, "Read and I will give you some nuts or figs…." With this stimulation the child tries to read. He does not work hard for the sake of reading itself, since he does not understand its value…. As his intelligence improves… his teacher may say to him, "Learn this passage or this chapter, I will give you a dinar or two."*
>
> *Now, all this is deplorable. However, it is unavoidable because of man's limited insight…. This is what the sages meant when they said, "A man ought always to labor in the Torah even if not for its own sake! For doing it not for its own sake, he may come to do it for its own sake"* (Pesachim 50b)
>
> — Maimonides, *Commentary on the Mishnah*
> (as translated by Isadore Tversky 1972, 404–07)

One of the most economically important educational milestones in many countries and in some US states is an exit exam for high school seniors, especially those intent on going to college. Examples include British A-levels, the French Baccalaureate, and the New York State Regents examinations. Since the 2001 No Child Left Behind Act, many US states have also adopted proficiency exams as a requirement for most high school graduates (Thomas S. Dee and Brian A. Jacob 2007). While American exit exams are not as closely linked to higher education as the

* Angrist: Department of Economics, Masssachusetts Institute of Technology, 50 Memorial Drive, Building E52, Room 353, Cambridge, MA 02142-1347, and NBER (e-mail: angrist@mit.edu); Lavy: Department of Economics, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel, NBER, and Department of Economics, Royal Holloway, University of London (e-mail: msvictor@huji.ac.il). Special thanks go to Katherine Eyal, Bruno Ferman, Rema Hanna, Alex Levkov, Nirupama Rao, Issi Romem, Simone Schaner, Yannay Spitzer, and Tamar Roth for outstanding research assistance in Cambridge and Jerusalem, to Dan McCaffrey for helpful discussions and for sharing his BRL code, to Joram Mayshar for helpful discussions, and to Esther Toledano and the staff of the research department at the Israeli National Insurance Institute for careful work and assistance analyzing postsecondary enrollment data. Thanks also go to Daron Acemoglu, Abhijit Banerjee, David Card, Mark Dynarksi, Sue Dynarksi, Ron Ehrenberg, Jinyong Hahn, Guido Imbens, Alan Krueger, Adriana Kugler, Kevin Lang, Thomas Lemieux, Doug Staiger, and seminar participants at UC-Berkeley, Boston University, Case Western University, CEMFI, Harvard University, MIT, Hebrew University, McMaster University, Princeton University, SOLE, Tel Aviv University, UCLA, RAND, University of Colorado, and Washington University for helpful discussions and comments. Finally, we thank the editor and referees for helpful comments. The 2001 Achievement Awards program was funded by the Israel Ministry of Education and administered by the division for secondary schools. We also gratefully acknowledge funding under NIH grant 1R01HD043809–01A1 and from the Falk Institute for Economic Research in Israel. The statements in the paper reflect the views of the authors and have not been endorsed by the program sponsors or funding agencies.

European exams, public university systems in the United States are increasingly likely to offer full tuition scholarships to high scorers.[1]

In Israel, the high school matriculation certificate, or Bagrut, awarded after a sequence of subject tests, is a formal prerequisite for university admission and arguably marks the dividing line between the working class and the middle class. Our estimates from the 1995 Israeli census suggest that even after controlling for highest grade completed, Bagrut holders earn 25 percent higher wages. Yet, in spite of the Bagrut's apparent economic and social value, Israeli society is marked by vast differences in Bagrut completion rates across regions and by socioeconomic background. As with high stakes exit exams in other countries, these disparities have led Israeli educators to implement remedial programs in an attempt to increase high school matriculation rates, with no apparent effect. These disappointing results echo similar findings from randomized trials in the US, where an array of service-oriented antidropout demonstrations for American teens have failed to increase high school graduation rates (Mark Dynarski and Philip Gleason 2002).

The discouraging results from previous antidropout interventions stimulated our interest in a simpler approach that focuses on immediate financial incentives for student effort. As a theoretical matter, cash incentives may be helpful if low-achieving students have high discount rates, reduce investment in schooling because of part-time work or other activities, or face peer pressure not to study. The promise of immediate financial rewards may tip the scales in favor of schoolwork. In this paper, we analyze the Achievement Awards demonstration, a project that provided cash awards for low-achieving high school students in Israel. The intervention discussed here rewarded Bagrut completion and performance on Bagrut subject tests with direct payments to students.

Though the intervention studied here is unusual in some ways, there is growing interest in student incentive programs in primary and secondary education. Most visible among these is a nascent effort involving achievement incentives across the New York City school system, including a $600 payment for each passing grade on New York State Regents exams.[2] In the same spirit, the Baltimore City Public School District has recently begun to pay a bonus of up to $110 to students who improve their scores on state graduation exams.[3] Our treatment is also closely related to the Texas Advanced Placement Incentive Program, which pays students (and their teachers) for success on high school Advanced Placement exams (see C. K. Jackson (2007) for a quasi-experimental evaluation). APIP student participants receive up to $500 for each AP exam passed.

Earlier demonstration projects involving student incentives include the Quantum Opportunities Program in US cities (Myles Maxfield, Allen Schirm, and Nuria Rodriguez-Planas 2003); the Learning, Earning, and Parenting (LEAP) demonstration project in Ohio (David M. Long, et al. 1996); Progresa in Mexico (Jere R. Behrman, P. Sengupta, and P. Todd 2000; Paul T. Schultz 2004); a program in Colombia (PACES) that provided private school vouchers (Angrist et al. 2002); and a recent randomized demonstration of a scholarship program for girls in Kenya (Michael Kremer, Edward Miguel, and Rebecca Thornton, forthcoming).[4] The Achievement

[1] For example, the University of Massachusetts offers free tuition to those with scores in the upper quartile of their school district's Massachusetts Comprehensive Assessment System (MCAS) score distribution.

[2] Jennifer Medina "Schools Plan to Pay Cash for Marks." *The New York Times,* June 19, 2007. http://www.nytimes.com/2007/06/19/nyregion/19schools.html.

[3] Katie Ash, "Promises of Money Meant to Heighten Student Motivation." *Education Week*, February 13, 2008. http://www.edweek.org/ew/articles/2008/02/13/23cash_ep.h27.html?qs=Promises_of_Money_Meant_to_Heighten_Student_Motivation.

[4] As far as we know, ours and the study by Kremer, Miguel, and Thornton (forthcoming) are the first completed projects using random assignment to evaluate substantial achievement-based payments to primary or secondary students using a randomized experimental design. Randomized trials investigating achievement incentives for college students include the Angrist, Daniel Lang, and Philip Oreopoulos (2009) evaluation of incentives and services for college freshman, an ongoing experiment involving community college students (Dan Bloom and Colleen Sommo 2005), and studies of Amsterdam college students by Edwin Leuven, Hessel Oosterbeek, and Bas van der Klaauw (2003) and Leuven et al. (2008).

Awards demonstration also has elements in common with the college tuition subsidy programs run by the I Have a Dream Foundation, and Robert Reich's proposal[5] to pay targeted bonuses of $25,000 to high school graduates from low-income families. Finally, any merit-based scholarship, such as the long-running but little-studied National Merit and National Achievement Scholarship Programs, has elements in common with Achievement Awards.[6]

On balance, the findings reported here point to an increase in Bagrut rates in treated schools, the result of sharply increased certification rates for girls. The overall effect on girls is on the order of 0.10 (as compared to a mean rate of about 0.29). Moreover, the increase for girls is driven by a group we think of as marginal, that is, girls with high predicted Bagrut rates *relative to other girls in our sample*. The prospects for this group are not rosy but they are not hopeless either. In particular, marginal students are those for whom certification is "within reach," in the sense that adjustments in test-taking strategy or study time are likely to have a payoff. We show, for example, that treated girls took only a few more tests but sharply increased their likelihood of meeting distribution requirements. Treated girls were also more likely to increase participation in a Spring study marathon. These findings point to more successful test-preparation and test-taking strategies.

The results naturally raise the question of whether the increased certification rate has a longer-term payoff since additional study time and changes in test-taking strategy need not generate additional human capital. Importantly, therefore, we show that treated girls in the marginal group were substantially more likely to enroll in higher education five years after the intervention. As far as we know, ours is the first study to document this sort of long-run benefit in response to achievement awards.

The rest of the paper is organized as follows. Section I sketches some of the theoretical background motivating our intervention. Section II describes program details and discusses data and descriptive statistics. Section III outlines the econometric framework. Section IV discusses the effects on Bagrut rates and some of the mechanisms and channels for these effects, including the differential response by gender. This analysis suggests that female Bagrut rates increased partly because treated girls were more likely to devote extra time to Bagrut preparation, while boys essentially ignored the program. Section V discusses the results for postsecondary schooling outcomes, and Section VI concludes.

I. Theoretical Context

Why do young men and women fail to complete high school? Why don't more go to college? These questions present something of a puzzle since the economic returns to schooling appear to be very high, and likely to exceed the costs of additional schooling for most non–college graduates. Research on education choices suggests possible explanations for low schooling levels, mostly related to heterogeneity in costs (or perceived costs) and heterogeneity in returns (or expected returns). Using data from the National Longitutional Surveys of Youth (NLSY), for example, Zvi Eckstein and K. I. Wolpin (1999) link drop-out decisions to lack of ability and motivation, low expectations about the rewards from graduation, disutility from schooling, and a comparative advantage in the jobs available to nongraduates. Another consideration raised in the literature on college attendance is liquidity constraints and the role of financial aid (see,

[5] Robert Reich, "Putting the Surplus, if Any, to Work." *The New York Times*, January 9, 1998. http://query.nytimes.com/gst/fullpage.html?res=9B06E4DF1F30F93AA35752C0A96E958260&scp=5&sq=robert%20B.%20reich%201998&st=cse.

[6] The National Merit programs give recognition and modest cash awards to a handful of high-achieving students based on their PSAT scores.

e.g., Winship C. Fuller, Charles F. Manski, and David A. Wise 1982). Liquidity constraints also surface in the literature on market work and the time students devote to their studies (e.g., John H. Tyler 2003; Todd R. Stinebrickner and Ralph Stinebrickner 2003). This literature notes that the causal links between student time allocation and educational outcomes are hard to quantify using observational data, though most of the evidence suggests student effort matters to at least some extent (Joseph V. Hotz et al. 2002; Stinebrickner and Stinebrickner 2004, 2008).

A number of features of the Israeli economic and social environment dovetail with the issues raised in previous research on low educational attainment. First, the fact that many Israeli students work, especially those from poorer families, suggests a link between low achievement and credit constraints. Paid employment may come at the expense of participation in widely available remedial programs that might make Bagrut success more likely.[7] A related concern is that some teenagers act as if they have very high discount rates (see, e.g., Jonathan Gruber 2001). Israeli requirements for compulsory military service (three years for boys and two years for girls) probably exacerbate the impact of discounting since working life for a male college graduate does not begin until six to seven years after high school. Uncertainty about returns may also be greater for poor Israelis, who are disproportionately likely to live in small towns with few educated adult role models. Finally, peer effects may be a negative influence in some of the relatively isolated communities where education levels are lowest.

The experimental program discussed here, which we refer to as the Achievement Awards demonstration, was motivated by a desire to tip the scales toward current investment in schooling and away from market work or leisure. Achievement Awards can be understood as a version of Reich's proposal to offer students from low-income families in the United States a $25,000 cash bonus for graduating high school. Michael P. Keane and Wolpin (2000) simulated the impact of the Reich policy in the context of a structural model of education choice. They estimated that this program would have a large impact on high school graduation rates and college attendance, especially for blacks. Although our research design cannot distinguish between alternative explanations for a failure to complete the Bagrut, almost any economic theory implies that financial incentives for completion should have at least a modest effect on some students (in particular, those close to the margin for success in the absence of treatment).[8] Our investigation also addresses the question of whether incentives might be counterproductive. Specifically, some psychologists have argued that financial or other rewards for student achievement replace powerful and enduring "intrinsic motivation," with short-lived "extrinsic motivation" that ultimately reduces achievement (see, e.g., Alfie Kohn 1999). If so, we should see negative program effects, at least in our long-term follow-up data on college enrollment.

An important feature of our investigation is a focus on students we see as on the margin for success, that is, those who are *relatively* close to Bagrut certification. Among students from the control schools in our sample, about 43 percent finished school with the minimum credit-units criterion for Bagrut status (20 units) satisfied, yet only 48 percent of these end up with a Bagrut, with others falling short because they are a couple units away or because of a failure to meet distribution requirements. For the purposes of our study, we defined marginal students by predicting Bagrut success in the control group using exams taken in tenth and eleventh grade. A key covariate in the prediction equation is the average test score on earlier exams. We also use this variable

[7] Roughly 36 percent of surveyed boys and 23 percent of surveyed girls in our sample reported working for pay in the last six months of the school year. Boys worked 31 hours per month and girls worked 13 hours per month (including zeros for nonworkers). Employment rates are higher for low-achieving boys than for relatively high achievers. As it turns out, however, the results discussed below suggest that changes in part-time employment are not what drive the effects we find on Bagrut rates.

[8] The idea that incentives motivate schoolchildren has a long history, as suggested by our epigraph from the twelfth-century biblical scholar Maimonides, who quotes tractate Pesachim from the Talmud.

to define marginal groups directly. The average probability of success in the marginal group so defined is close to 37 percent for boys and 52 percent for girls, in contrast with only a 3–6 percent success rate for nonmarginal students. The substantial cross-sectional and time-series variation in Bagrut rates suggests that, under some circumstances, students in this marginal group can clear the final Bagrut hurdle.[9]

II. Program Details

A. *Bagrut in the Israeli School System*

High school students earn a Bagrut by passing a series of national exams in core and elective subjects beginning in tenth grade, with more tests taken in eleventh grade and most taken in twelfth grade. Thus, Bagrut certificates are typically obtained at the end of senior year or later. Students choose to be tested at various proficiency levels, with each test awarding one to five credit units per subject, depending on difficulty. Some subjects are mandatory and many must be taken for at least three units. A minimum of 20 credit units is required to qualify for a matriculation certificate, though some study programs require more, and students must also satisfy distribution requirements. About 52 percent of all high school seniors received a matriculation certificate in the 1999 and 2000 cohorts (Israel Ministry of Education 2001). Roughly 60 percent of those who took at least one Bagrut subject test ended up receiving a Bagrut certificate. In our sample, however, Bagrut rates are much lower.

Although the Bagrut is an Israeli institution, it can be understood in the American vernacular as a "college-bound" indicator. Most of the Israeli students who fail to complete a Bagrut still finish their secondary schooling. Nevertheless, postsecondary schooling options for high school graduates without a Bagrut are limited; very few will obtain further schooling. Even institutions that are not otherwise very selective, such as teachers' colleges and two-year professional programs for nursing, optometry, and computer programming, favor applicants with a Bagrut. Consistent with this, regression evidence from the Israeli census suggests that the economic returns to a Bagrut are high, though we do not have a good experiment for the earnings consequences of Bagrut certification. Clearly, however, if certification increases schooling, it very likely increases earnings. A recent quasi-experimental study of exit exams in Texas suggests that those who pass these exams go on to get more postsecondary schooling than they otherwise would have (Francisco Martorell 2005).

B. *Research Design and Program Implementation*

In December 2000, Ministry of Education officials selected the 40 nonvocational high schools with the lowest 1999 Bagrut rates in a national ranking, but above a minimum threshold rate of 3 percent. Some low-rate schools in the relevant universe were ineligible to participate in the experiment for technical or administrative reasons. We also felt that schools with virtually no Bagrut recipients were unlikely to benefit from the program. The list of participants includes 10

[9] Note that our use of the notion of a marginal student differs from that in the literature on instrumental variables (IV) estimates of the returns to schooling, where IV estimates are driven by individuals whose schooling decisions can be affected by a policy change such as a change in compulsory attendance laws (see, e.g., Kevin Lang 1993; David E. Card 1995; James J. Heckman and Edward Vytlacil 1998). Here we use the term marginal descriptively: students in our marginal group are those relatively close to success. This also turns out to be the group most affected by treatment, but we do not explore the question of whether the economic returns to schooling for this group differ from population average returns, as might be the case if credit constraints are important.

Arab and 10 Jewish religious schools and 20 Jewish secular schools.[10] The total number of treated schools was determined by the program budget constraint. Ultimately, about $650,000 (3.1 million shekels) was awarded. Treatment was randomly assigned to 20 of the 40 participating schools in December of 2000 and treated schools were contacted shortly after random assignment. Only schools selected for treatment were informed about the program, though data were collected for students attending all schools in the original group of 40 (except for one control school that closed). As far as we know, no control school administrators ever inquired about the program.

Although not large enough to ensure perfect treatment-control balance, the number of clusters used here is typical of other group-randomized trials (GRTs) (see, e.g., Ziding Feng et al. 2001 or Allan K. Donner, Stephen Brown, and Penny Brasher 1990). To improve treatment-control balance, the assignment used a matching strategy that paired treatment and control schools based on lagged values of the primary outcome of interest, the average 1999 Bagrut rate. (Bagrut rates from 1999 were used to select and match schools within pairs because the 2000 data were incomplete when treatment was assigned.) Treatment was assigned randomly within pairs, the most common matching strategy in GRTs (see, e.g., Mitchell H. Gail et al. 1996).

Appendix Table A1 reports Bagrut means and enrollment counts for 1999–2002 in each of the 39 schools that participated in the study (the control school in pair 6 had closed by the time treatment was assigned). The 1999 Bagrut rates used for matching ranged from 3.6 to 28.6 percent and are (by design) similar within pairs. Schools ranged in size from 10 to 242 seniors in 1999, and some schools show marked changes in size from year to year. These changes reflect the unstable environment that characterizes Israel's weakest schools. For example, some absorb cohorts of new immigrants.[11]

The (student-weighted) treatment-control difference in 1999 Bagrut rates is only about -0.018 (s.e. $= 0.054$), not surprisingly since 1999 rates were used to construct pairs. Bagrut rates in 2000 were not as well-matched, though the 2000 treatment-control difference is not significantly different from zero. The overall difference in 2000 rates is 0.048 (s.e. $= 0.055$), and the difference for girls is 0.083 (s.e. $= 0.063$). These comparisons and similar balancing tests for covariates are reported in Appendix Table A2. This table shows no significant treatment-control differences in covariates in either the pretreatment or treatment years. For example, the pretreatment difference in father's schooling in the 2000 sample is 0.328 (s.e. $= 0.716$) and the treatment-year difference in lagged test scores is 1.17 (s.e. $= 4.51$), both insignificant and modest relative to the standard deviation of these variables. At the same time, while not significantly different from zero, pretreatment differences in Bagrut rates, especially for girls, are a source of concern. This leads us to estimate models that combine data for 2000 and 2001 and include school fixed effects.

Program Parameters.—Every student in treated schools who received a Bagrut was eligible for a payment. Randomized trials that assign treatment status to entire schools are often more attractive than within-school randomization of individual students for both programmatic and logistical reasons. First, school-based assignment reduces the perception of unfairness that may be associated with randomization. Second, students not offered treatment may nevertheless be

[10] Israel runs separate school systems for secular Jews, religious Jews, and Arabs. Rules and standards for Bagrut certification are similar in all three systems.

[11] The variability in Bagrut rates in later years results from small school size, changes in school populations due to immigration and internal migration, and measurement error in the Bagrut data. In practice, the 1999 Bagrut rate is not as powerful a predictor of the 2000 and 2001 Bagrut rates as we had hoped, although it is still worth something. The R^2 from a weighted regression of the 2001 rate on the 1999 rate is 0.15. It bears emphasizing that substantial variability in year-to-year performance measures for individual schools is not unique to our sample. Kane and Staiger (2002) report that much of the year-to-year variation in school performance in North Carolina is due to school-level random shocks that come from sources other than sampling variance.

affected by the treatment received by other students in the same school, diluting within-school treatment effects. Finally, successful implementation of educational interventions depends partly on the cooperation of teachers and school administrators, and the intervention may get additional leverage from peer effects when classmates participate.

The timing of the program and key data collection points are summarized in a chart in the Appendix. The orientation for principals was in January 2001, about one-third of the way into the 2000–2001 school year. Follow-up contacts in March 2001 were used to verify participation by contacting principals and school administrators. Five treatment schools are noncompliers in the sense that, following the orientation session in January 2001, principals in these schools had taken no concrete actions to inform pupils or teachers about the program and/or indicated that they did not wish to participate.

The Achievement Awards program was meant to last three years, with awards given to high school students in every grade. Seniors became aware of the program about one-third of the way into the year, before a "big push" Bagrut study effort that is traditional in the spring. Modest awards were offered to students who progressed from tenth to eleventh grade and from eleventh to twelfth grade. Small awards of NIS500 were also given for taking any Bagrut component test, regardless of the outcome, with NIS1,500 to be given for passing component tests before senior year. The largest award was NIS6,000 (almost $1,500) for any senior who received a Bagrut. The total amount at stake for a student who passed all achievement milestones was NIS10,000 or just under $2,400. This is about one-third of the after-tax annual earnings a student could expect from working full-time as a high school dropout, and about twice as much as a student might earn working full time in two summer months.[12]

In practice, program implementation was affected by a number of events. First, following an election and a change of government, a new minister of education was appointed. Second, a budgetary crisis in late fall of 2001, the beginning of the second year of the program, led to a sharp reduction in the education budget. Because of these events and some media attention when the program became public knowledge, the award scheme for younger cohorts was eventually canceled. As a consequence, awards were given for only one year of achievement and the maximum amount awarded was NIS6,000. Although tenth and eleventh graders were eligible for more modest short-term awards, by the time the bulk of their Bagrut effort took place, the program had ceased. This disruption notwithstanding, the program for high school seniors operated as planned, and the post-intervention survey shows that most students were aware of specific program features. The analysis in this paper is therefore limited to high school seniors.

C. *Data and Descriptive Statistics*

Baseline data were collected in January 2001, while the main Bagrut outcome for the treated cohort comes from tests taken in June of 2001. A small follow-up survey was conducted in late summer and early fall of 2001. Students had an unanticipated opportunity to be tested again in August–September of 2001 and a regularly scheduled second chance in the winter of 2002. Though the results using winter 2002 data are similar, we prefer the June data because of the disruption and uncertainty introduced by the Bagrut retests, which were unexpectedly offered in the fall of 2001.

In addition to Bagrut outcomes, our administrative dataset includes basic socioeconomic information. This information is summarized in Table 1, which presents 2000–2001 means for our sample of schools and the nation. By construction, the Bagrut rate in the experimental

[12] These estimates are based on the minimum wage rate for the 17–18 age group in 2001, taken from a company that compiles these data over time (see http://www.hilan.co.il/calc/MinimumWageCalculator.aspx).

TABLE 1—DESCRIPTIVE STATISTICS

	Experimental sample			National		
	All (1)	Boys (2)	Girls (3)	All (4)	Boys (5)	Girls (6)
Panel A. 2001						
Bagrut rate	0.243	0.200	0.287	0.629	0.574	0.678
School covariates						
Arab school	0.348	0.374	0.320	0.163	0.159	0.167
Religious school	0.115	0.084	0.148	0.170	0.154	0.184
Micro covariates						
Father's education	10.1	9.82	10.3	12.2	12.2	12.1
	(3.07)	(3.11)	(3.00)	(3.48)	(3.48)	(3.48)
Mother's education	10.0	9.87	10.2	12.0	12.0	11.9
	(3.29)	(3.32)	(3.24)	(3.42)	(3.42)	(3.42)
Number of siblings	3.74	3.65	3.84	2.97	2.91	3.03
	(2.66)	(2.64)	(2.68)	(1.95)	(1.91)	(1.98)
Immigrant	0.064	0.029	0.100	0.023	0.021	0.025
Lagged score	53.1	52.1	54.2	—	—	—
	(29.4)	(29.4)	(29.3)			
Proportion missing						
Father's education	0.144	0.168	0.118	0.124	0.128	0.120
Mother's education	0.153	0.173	0.132	0.136	0.142	0.130
Number of siblings	0.116	0.111	0.122	0.107	0.110	0.105
Observations	3,821	1,960	1,861	76,990	36,423	40,567
Panel B. 2000						
Bagrut rate	0.224	0.177	0.272	0.611	0.560	0.657
School covariates						
Arab school	0.319	0.352	0.286	0.161	0.160	0.163
Religious school	0.134	0.098	0.170	0.171	0.154	0.186
Micro covariates						
Father's education	9.87	9.75	10.0	12.1	12.1	12.0
	(3.07)	(3.15)	(2.99)	(3.56)	(3.57)	(3.56)
Mother's education	9.80	9.71	9.9	11.9	11.9	11.9
	(3.26)	(3.33)	(3.18)	(3.48)	(3.50)	(3.45)
Number of siblings	3.68	3.53	3.84	2.99	2.92	3.06
	(2.47)	(2.34)	(2.58)	(1.98)	(1.92)	(2.03)
Immigrant	0.074	0.039	0.109	0.032	0.029	0.035
Lagged score	50.2	49.1	51.4	—	—	—
	(28.9)	(29.4)	(28.4)			
Proportion missing						
Father's education	0.109	0.121	0.096	0.087	0.094	0.080
Mother's education	0.115	0.129	0.100	0.085	0.094	0.077
Number of siblings	0.101	0.105	0.098	0.103	0.107	0.100
Observations	4,039	2,038	2,001	77,241	36,484	40,757

Notes: Columns 1–3 report sample means. Standard deviations are shown in parentheses. Statistics in columns 4–6 are from the authors' tabulation of administrative data for schools with a positive Bagrut rate in 1999.

sample is low; 22–24 percent versus 61–63 percent nationally (among schools with a positive Bagrut rate in 1999). Relative to the national average, the experimental sample includes more students attending Arab schools, but fewer attending religious schools. Students in the experimental sample also have less educated parents, more siblings, and are more likely to be new immigrants than in the country as a whole.[13]

III. Econometric Framework

The following model is used to estimate treatment effects using data for individual students:

$$y_{ij} = \Lambda[\mathbf{X}_j'\boldsymbol{\alpha} + \textstyle\sum_q d_{qi}\delta_q + \mathbf{W}_i'\boldsymbol{\beta} + \gamma z_j] + \varepsilon_{ij}, \tag{1}$$

where i indexes students and j indexes schools, and $\Lambda[\cdot]$ is a possibly nonlinear link function (in this case, the logistic transformation or the identity). We assume that $E[y_{ij}] = \Lambda[\mathbf{X}_j'\boldsymbol{\alpha} + \sum_q d_{qi}\delta_q + \mathbf{W}_i'\boldsymbol{\beta} + \gamma z_j]$, where the expectation is conditional on individual and school characteristics (alternately, this is the minimum mean square error approximation to the relevant conditional expectation). School-level variables include the treatment dummy, z_j, and a vector of school-level controls, $\mathbf{X}_j$, that includes a dummy for Arab schools and a dummy for Jewish religious schools. In some specifications, this vector also includes dummies for randomization pairs. The vector $\mathbf{W}_i$ contains individual characteristics such as parental schooling, the number of siblings, and immigrant status. Some models also include a function of lagged test scores which, as we show below, predicts Bagrut status exceptionally well. This function consists of three dummies $\{d_{qi}; q = 2, 3, 4\}$ indicating the quartile of a student's credit-unit-weighted average test score on tests taken before January 2001, when the program was implemented.[14] We also estimated models replacing score-group dummies with a linear term in lagged scores.

The first econometric issue raised by our study is noncompliance. In follow-up contacts in March 2001, we verified the level of compliance by contacting the 39 participating principals and school administrators (one school had closed). The principals of three noncompliant schools had taken no concrete actions to inform students or teachers about the program and/or indicated that they did not wish to participate. School administrators in two other schools designated as noncompliant hoped to participate but submitted student rosters shortly after the deadline. Because the decision to cooperate may be related to potential outcomes, even within pairs, we analyze the data based on randomly assigned intention to treat, i.e., the reduced-form impact of the randomized offer of program participation, indicated by z_j in equation (1). The conclusion briefly discusses the impact of adjustments for noncompliance. Because the compliance rate was high, this involves a modest rescaling of the reduced-form estimates.

A second statistical issue is how best to adjust inferential procedures for clustering at the school level. The traditional cluster adjustment relies on a linear model with random effects, an approach known to economists through the work of Brent Moulton (1986). When the clusters are all of size n, this amounts to multiplying standard errors by a "design effect," $[1 + (n - 1)\rho]^{1/2}$, where ρ measures the intra-cluster residual correlation. A problem with random effects models in this context is that the equi-correlated error structure they impose is implausible for binary outcomes like Bagrut status. Another problem is that estimates of ρ are biased and tend to be too low, making parametric cluster adjustments too optimistic (Mark D. Thornquist, and Garnet L. Anderson 1992; Feng et al. 2001).

[13] About 10–15 percent of the administrative records are missing socioeconomic characteristics. We imputed missing data using means by sex and school type. However, data on our core outcome variables, Bagrut status and postsecondary schooling, are essentially complete.

[14] In particular, we calculated each student's credit-unit-weighted average score as of January (coding zeros for those with no tests) and then divided students into quartiles on the basis of this weighted average. Students were assigned to quartiles using the distribution of credit-unit weighted average scores for their cohort in our sample.

A modern variation on random effects models is the Generalized Estimating Equation (GEE) framework developed by Kung-yee Liang and Scott L. Zeger (1986). GEE allows for an unrestricted correlation structure and can be used for binary outcomes and nonlinear models such as logit. The advantages of GEE are flexibility and availability in proprietary software. The primary disadvantage is that the validity of GEE inference turns on an asymptotic argument based on the number of clusters (as do parametric random effects models). GRTs often have too few clusters for asymptotic formulas to provide an acceptably accurate approximation to the finite-sample sampling distribution. As with parametric Moulton-type or design-effect adjustments, GEE standard errors are also biased downward (See, e.g., Jeffery M. Wooldridge 2003).

To sidestep the problem of downward-biased GEE standard errors, we estimated standard errors (for models without school effects) using Robert M. Bell and Daniel F. McCaffrey's (2002) Biased Reduced Linearization (BRL) estimator. BRL implements a correction for GEE standard errors similar to James G. MacKinnon and Halbert White's (1985) bias-corrected heteroskedasticity-consistent covariance matrix. Bell and McCaffrey present Monte Carlo evidence suggesting BRL generates statistical tests of the correct size in traditional random effects models with normally distributed errors. Appendix Table A3 compares standard errors from linear probability models similar to equation (1), estimated using alternative cluster adjustments. This table shows that the BRL standard errors, while slightly larger than those produced by the conventional GEE cluster adjustment, are similar to those arising from a two-step procedure based on adjusted group means proposed by Stephen Donald and Kevin Lang (2007). Since an analysis of data grouped at the cluster level is likely to be conservative, this similarity is encouraging. Web Table A3 also serves as a robustness check in the sense that the basic findings are indeed apparent in a school-level analysis of group means.

A final statistical issue is the role of pair effects, which we sometimes include in the vector of school-level controls, $\mathbf{X}_j$. Control for pair effects can be motivated by the fact that a school's pair identity is a stratification variable fixed at the time of random assignment. On the other hand, pair effects can be dropped without biasing estimates of treatment effects, since intention to treat is a (fair) coin toss in each pair. Moreover, ignoring stratification variables may lead to more precise estimates when these effects explain little of the variation in the dependent variable (Paula H. Diehr et al. 1995; Angrist and Jinyong Hahn 2004). Another practical consideration in this regard is that when estimating models with pair effects in separate samples of boys and girls, we lose some single-sex schools. We therefore report pooled-sample results from models with and without pair effects, while the single-sex models omit them.

IV. Results

A. *Cross-Section Estimates*

Estimates of equation (1) support the notion that the Achievement Awards program increased Bagrut rates in 2001, though there is also some evidence of spurious effects in the previous (2000) cohort. These findings can be seen in Table 2, which reports OLS and Logit estimates of equation (1) using the full sample, as well as separate results for boys and girls.[15] Panel A reports results for 2001, the posttreatment year, while panels B and C report results for the 2000 and 2002 cohorts as a specification check. The first set of results shown in each panel is from models that include school covariates; the second is from models that add lagged score quartile dummies and individual student characteristics (mother and father's schooling, the number of siblings, and

[15] The logit results are reported as marginal effects on the treated. The sample used for logit drops schools with zero dependent variable means.

Table 2—Treatment Effects and Specification Checks

	Pair effects	Boys + girls OLS (1)	Boys + girls Logit (2)	Boys OLS (3)	Boys Logit (4)	Girls OLS (5)	Girls Logit (6)
Panel A. 2001							
Dependent variable mean		0.243		0.200		0.287	
Model with:							
School covariates	No	0.056	0.051	−0.010	−0.011	0.105	0.093
		(0.049)	(0.045)	(0.052)	(0.055)	(0.061)	(0.053)
	Yes	0.052	0.054	—	—	—	—
		(0.047)	(0.043)				
School covariates, quartile dummies, micro covariates	No	0.052	0.047	−0.022	−0.023	0.105	0.097
		(0.039)	(0.039)	(0.043)	(0.045)	(0.047)	(0.046)
	Yes	0.067	0.055	—	—	—	—
		(0.036)	(0.036)				
Number of students		3,821		1,960		1,861	
Number of schools		39		34		34	
Panel B. 2000							
Dependent variable mean		0.224		0.177		0.272	
Model with:							
School covariates	No	0.050	0.046	0.045	0.040	0.075	0.069
		(0.056)	(0.051)	(0.060)	(0.055)	(0.067)	(0.061)
	Yes	0.043	0.045	—	—	—	—
		(0.059)	(0.058)				
School covariates, quartile dummies, micro covariates	No	0.030	0.018	0.009	0.006	0.066	0.051
		(0.041)	(0.042)	(0.050)	(0.052)	(0.046)	(0.046)
	Yes	0.043	0.030	—	—	—	—
		(0.044)	(0.046)				
Number of students		4,039		2,038		2,001	
Number of schools		39		33		35	
Panel C. 2002							
Dependent variable mean		0.305		0.257		0.357	
Model with:							
School covariates	No	−0.019	−0.019	−0.026	−0.028	−0.010	−0.010
		(0.071)	(0.071)	(0.073)	(0.075)	(0.077)	(0.078)
	Yes	−0.018	−0.018	—	—	—	—
		(0.050)	(0.059)				
School covariates, quartile dummies, micro covariates	No	−0.023	−0.021	−0.026	−0.024	−0.015	−0.014
		(0.044)	(0.045)	(0.046)	(0.047)	(0.046)	(0.046)
	Yes	−0.027	−0.033	—	—	—	—
		(0.033)	(0.034)				
Number of students		4,328		2,269		2,059	
Number of schools		38		33		33	

Notes: The table reports OLS estimates and logit marginal effects. Panel A shows treatment effects. Results from 2000 and 2002 are specification checks. BRL standard errors are reported in parentheses. Pair effects are omitted from models estimated separately for boys and girls so as not to lose pairs that include single-sex (religious) schools.

student's immigrant status). Each of these specifications was estimated with and without pair effects as additional controls.

Estimates from all models for the combined sample of boys and girls, reported in columns 1–2, are positive, though only those from models with a full set of controls and pair effects are (marginally) significant. For example, the OLS estimate in column 1 (with school controls) is

0.056 (s.e. = 0.049). The OLS estimate with school controls, lagged score quartile dummies, individual controls, and pair effects is larger and more precise, at 0.067 (s.e. = 0.036). The corresponding logit marginal effects, reported in column 2, are slightly smaller. There is some evidence of school-level random effects, however, since the estimates for 2000 are also positive. The largest of the estimates for 2000, in the first row of panel B, is almost as large as the corresponding estimate for 2001. On the other hand, the 2000 effects shrink (and hence the gap between the 2001 and 2000 estimates increases) when additional control variables are included in both models, and none of the estimates for 2000 is significantly different from zero. These results suggest that the treatment-control difference in Bagrut rates in 2000 is at least in part explained by a chance association between treatment status and student characteristics in that year.

Separate analyses by sex show sharp differences in effects for boys and girls. The estimates for boys, reported in columns 3–4, are uniformly small and negative; none is significantly different from zero, while the estimates for boys in 2000 are small, positive, and also insignificant. For example, 2001 estimates from models with all controls are −0.022 (OLS) and −0.023 (logit). In contrast, the 2001 estimates for girls are on the order of 0.10, and most are at least marginally significantly different from zero. Moreover, while the 2000 estimates for girls are also positive, none of these is as large as those for 2001, and all are insignificant.

It is also worth noting that the basic pattern of 2001 and 2000 results in Table 2 is apparent in an analysis of school-level grouped means, reported in Appendix Table A3. For example, the 2001 weighted grouped-data estimate for girls from a model with school covariates (adjusted for lagged test scores using the Donald and Lang (2007) two-step procedure) is 0.095 (s.e. = 0.044). Grouped-data estimates for girls without weighting are similar.

A causal interpretation of the 2001 effects for girls is further reinforced by the analysis of data from the 2002 graduating cohort. This can be seen in panel C of Table 2, which shows that Bagrut rates were remarkably similar in treatment and control schools in the year after the experiment ended.[16] The estimated treatment-control differences are on the order of −0.02, for both boys and girls. Although seniors in the 2002 cohort were offered small payment to take and pass at least one Bagrut subject test as eleventh graders in 2001, no further incentives were offered to this cohort since the program was cancelled before they began their senior year. Moreover, we found no evidence that the modest payments offered to eleventh graders affected their test-taking behavior or results. Thus, the treatment experienced by the 2002 cohort can be seen as providing a sort of placebo control in that these students attended treated schools, but were exposed to little in the way of a changed environment.

The estimates in Table 2 suggest that the Achievement Awards program had no effect on boys, but show reasonably clear evidence of increased Bagrut rates for girls. The picture is muddied somewhat by the positive effects on girls in 2000, though the 2002 results suggest that the imbalance in 2000 was transitory. Nevertheless, in an effort to reduce any possible bias from school-level omitted variables, we estimated models using stacked 2000, 2001, and 2002 data in a setup that includes school effects. This procedure controls for any time-invariant school-level omitted variable that might explain higher Bagrut rates in treated schools. The introduction of school (fixed) effects also provides an alternative approach to the clustering problem. Before turning to the models with school fixed effects, however, we refine the estimation strategy by isolating the group of marginal students most likely to benefit from the Achievement Awards intervention. We first report results for marginal groups using models for levels, and then turn to models for marginal groups incorporating school effects.

[16] The 2002 Bagrut levels are higher than those for 2000–2001 because the 2002 data reflect additional attempts at certification after the main testing round for high school seniors. This difference should not affect the comparison of treatment and control groups.

B. *Identification of Marginal Students*

We identified groups of marginal students using a predictive regression that models the probability of success as a function of school characteristics and individual covariates. The predictive model is

$$(2) \qquad y_{ij} = \Lambda\,[\mathbf{X}_j'\boldsymbol{\pi}_0 + \textstyle\sum_q d_{qi}\pi_q + \mathbf{W}_i'\boldsymbol{\pi}_1] + \eta_{ij},$$

where $\mathbf{X}_j$, $\mathbf{W}_i$, and $\{d_{qi};\ q = 2, 3, 4\}$ are as in equation (1). This model is solely a classification device, so we estimated it using the 2001 control sample only. Some of the specifications omit mother's schooling from $\mathbf{W}_i$ since the parents' schooling effects are never both significantly different from zero.

The Logit coefficients estimated using equation (2), reported in Table 3, show that the three lagged score quartile dummies are far and away the best predictor of Bagrut status. Especially noteworthy is the fact that the lagged score coefficients dwarf family background effects, both in size and statistical significance.[17] This is not surprising, since the marginal probability of Bagrut certification in June 2001 was about 1 percent in the lowest score quartile, 9 percent in the second score quartile, 29 percent in the third score quartile, and 49 percent in the upper quartile. This gradient reflects the fact that Bagrut status is determined in large part by accumulating credit units. Students entering senior year with very few units simply cannot make up the shortfall. On the other hand, the odds of Bagrut success are substantial for students who've done well on tenth and eleventh grade subject tests. Conditional on lagged scores, family background is of modest value as a predictor of Bagrut status, though background covariates do more for girls than for boys.

Motivated by these results, we reestimated equation (1) using two subgroup classification schemes. The first splits students according to the lagged score distribution, again using credit unit–weighted scores. In other words, we divided students into two roughly equal-size groups, the top half with $d_{3i} = d_{4i} = 1$. Second, we used the fitted values from model (2), again dividing students into roughly equal-size groups. This second scheme provides a check on the notion that high lagged scores identify students who have a shot at certification. For both schemes, we estimated models that include school covariates and either a quartile dummy calculated from the distribution of the classification variable, or a linear term in the classification variable.[18] Both classification schemes appear to do a good job of isolating students likely to be affected by treatment, a fact documented in the descriptive statistics at the top of Table 4. In particular, both divide the sample into a low-achieving group with a 3–6 percent chance of certification and a relatively high-achieving group that faces much better odds. Girls in the top group have a better than 50 percent probability of Bagrut success.

Panel A of Table 4, which reports treatment effects in 2001, consistently shows small and insignificant estimates for both boys and girls in the bottom half of the distribution of lagged scores or fitted values from equation (2). In contrast, the estimates for girls in the top half of the distribution, reported in columns 3 and 7, show large significant effects. For example, the estimated effect on girls in the top half of the distribution of lagged scores is 0.206 (s.e. = 0.079), while the corresponding effect on girls in the top half of the distribution of fitted values is 0.194 (s.e. = 0.077). Moreover, the corresponding estimates for girls in 2000, reported in panel B, are

[17] Here we show logit coefficients instead of marginal effects since it is the *relative* predictive power of covariates that is of primary interest.

[18] The results are almost identical when a scaled score is substituted for the credit-unit lagged raw score used to define classification groups. Students were divided into lagged-score or predicted-Bagrut groups based on models or score distributions estimated within gender and year.

Table 3—Determinants of Bagrut Status in the Control Group

	Boys + girls		Boys		Girls	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable mean	0.219		0.194		0.239	
School covariates						
Arab school	0.987	1.08	0.625	0.901	1.30	1.26
	(0.540)	(0.558)	(0.830)	(0.905)	(0.420)	(0.400)
Religious school	0.627	0.632	2.56	2.57	0.282	0.281
	(0.579)	(0.578)	(0.603)	(0.628)	(0.508)	(0.508)
Micro covariates						
Father's education	0.076	0.056	0.081	0.016	0.068	0.074
	(0.044)	(0.040)	(0.053)	(0.047)	(0.039)	(0.038)
Mother's education	—	0.034	—	0.109	—	−0.012
		(0.037)		(0.079)		(0.032)
Has more than 4 siblings	−0.431	−0.422	0.080	0.093	−0.682	−0.685
	(0.307)	(0.308)	(0.521)	(0.547)	(0.254)	(0.255)
Immigrant	0.924	0.905	1.31	1.35	1.10	1.12
	(0.383)	(0.385)	(1.204)	(1.196)	(0.364)	(0.364)
Lagged score quartile dummies						
2nd	2.83	2.84	1.77	1.78	3.16	3.16
	(0.739)	(0.738)	(0.871)	(0.880)	(0.784)	(0.785)
3rd	4.34	4.35	3.81	3.86	4.69	4.69
	(0.662)	(0.664)	(0.788)	(0.776)	(0.781)	(0.781)
4th	5.02	5.01	4.57	4.56	5.33	5.33
	(0.581)	(0.583)	(0.629)	(0.625)	(0.738)	(0.735)
Number of students	1,876		850		1,026	
Number of schools	19		15		18	

Notes: The table reports logit estimates. The estimates in this table were constructed using the sample of 2001 control schools only. BRL standard errors are reported in parentheses.

less than half as large and none is significantly different from zero. Especially encouraging is the fact that the estimates using 2002 data, reported in panel C, are essentially zero for both boys and girls, in both the top and bottom half of the lagged-score or predicted-success distribution.

Models allowing interactions with individual demographic characteristics generated inconclusive results. We were unable to find a subgroup of boys exhibiting a pattern of strong treatment effects similar to that for girls. We also looked for interactions with parental education both in the full sample and for marginal groups. These models included an interaction of the treatment variable with parents' schooling measured in years or using an indicator for students whose parents had at least 12 years of schooling. This generated no significant interactions, a finding that weighs against credit constraints as the primary force responsible for low student achievement in our sample.

C. *Models with School Effects*

Table 4 shows significant treatment effects for girls in the upper half of the 2001 lagged score distribution. On the other hand, as noted above, the presence of some fairly large (though insignificant) treatment-control differences in 2000 raises a concern about omitted school effects. We therefore estimated stacked models for multiple years controlling for additive school fixed effects. The coefficient of interest in the stacked specification is the interaction between a dummy for

Table 4—Estimates in Covariate Subgroups

	By lagged score				By predicted probability			
	Boys		Girls		Boys		Girls	
	Top (1)	Bottom (2)	Top (3)	Bottom (4)	Top (5)	Bottom (6)	Top (7)	Bottom (8)
Panel A. 2001								
Dependent variable mean	0.365	0.035	0.518	0.056	0.368	0.032	0.518	0.056
Models with:								
School covariates, quartile dummies	−0.013	0.007	0.206	−0.020	−0.047	0.005	0.194	−0.015
	(0.083)	(0.016)	(0.079)	(0.024)	(0.077)	(0.016)	(0.077)	(0.023)
School covariates, linear lagged score or predicted prob.	−0.009	0.007	0.213	−0.021	−0.044	0.001	0.207	−0.019
	(0.083)	(0.017)	(0.079)	(0.022)	(0.079)	(0.017)	(0.078)	(0.026)
Number of students	980	980	933	928	980	980	932	929
Panel B. 2000								
Dependent variable mean	0.318	0.035	0.475	0.068	0.320	0.033	0.478	0.066
Models with:								
School covariates, quartile dummies	0.055	−0.014	0.098	0.009	0.033	0.004	0.086	0.009
	(0.079)	(0.035)	(0.074)	(0.027)	(0.078)	(0.027)	(0.071)	(0.023)
School covariates, linear lagged score or predicted prob.	0.055	−0.014	0.094	0.007	0.010	0.000	0.089	0.007
	(0.079)	(0.035)	(0.072)	(0.026)	(0.077)	(0.028)	(0.070)	(0.024)
Number of students	1,022	1,016	1,004	997	1,021	1,017	1,002	999
Panel C. 2002								
Dependent variable mean	0.475	0.040	0.611	0.101	0.472	0.042	0.608	0.106
Models with:								
School covariates, quartile dummies	−0.018	−0.004	−0.017	−0.030	−0.029	−0.007	−0.006	−0.021
	(0.101)	(0.016)	(0.088)	(0.032)	(0.098)	(0.017)	(0.078)	(0.029)
School covariates, linear lagged score or predicted prob.	−0.008	−0.003	−0.013	−0.037	−0.032	−0.015	−0.001	−0.020
	(0.097)	(0.016)	(0.088)	(0.031)	(0.088)	(0.021)	(0.073)	(0.028)
Number of students	1,135	1,134	1,035	1,024	1,135	1,134	1,030	1,029

Notes: The table reports logit marginal effects in top and bottom subgroups, classified by lagged test scores or predicted probability of Bagrut success (as a function of lagged scores and covariates). Panel A shows treatment effects. Results from 2000 and 2002 are specification checks. BRL standard errors are reported in parentheses.

2001 and the treatment indicator, z_j. The resulting estimates can be somewhat loosely interpreted as a student-weighted difference-in-differences procedure comparing treatment effects across years.[19] Models with school effects control for time-invariant omitted variables, a particular concern given the positive estimates in 2000 data. Moreover, school effects provide an alternative control for school-level clustering and absorb some of the variability in average Bagrut rates by school, possibly leading to a gain in precision.[20]

[19] The differences-in-differences analogy is imperfect for two reasons. First, the estimates are implicitly weighted by the number of students in each school. Second, the panel is unbalanced across years because a few schools that were open in one year were closed in another. In addition, any school with a mean Bagrut rate of zero drops out of the logit sample.

[20] Note that BRL standard errors adjust *inference* for the uncertainty generated by omitted random effects, while models with school fixed effects change the estimator.

The equation used to estimate logit models with school effects can be written:

$$y_{ijt} = \Lambda\,[\mu_j + \xi_t + \mathbf{X}_j'\boldsymbol{\alpha}_t + \mathbf{W}_i'\boldsymbol{\beta} + \gamma(z_j d_t)] + \varepsilon_{ijt}, \tag{3}$$

where $t = 2000$–2002; $d_t = 1[t = 2001]$; μ_j is a school effect and ξ_t is a year effect; and α_t is a year-specific vector of coefficients on school covariates. The micro covariates, $\mathbf{W}_i$, now include either a dummy for first quartile students (in the bottom half) or third quartile students (in the top half) or a linear term in lagged score or predicted Bagrut success, depending on how the sample was divided into high and low achievers. The dependent variable, y_{ijt}, is the Bagrut status of student i in school j in year t.

Not surprisingly given the baseline differences in 2000, estimates of the model with school effects using 2001 and 2000 data are smaller than the corresponding estimates for 2001 only. This can be seen in panel A of Table 5. The estimated matriculation gains for girls in the upper half of the lagged score distribution are about 0.093 (s.e. = 0.043) in a model with a dummy for third-quartile students (column 3), and 0.102 (s.e. = 0.043) in a model with linear control for the lagged score. The corresponding estimates when students are split by predicted Bagrut rates (column 7) are 0.08 and 0.09. The estimated effects on girls in the bottom half of the distribution of the classification variable are small and insignificantly different from zero. The estimates for boys are close to zero in all subgroups.

Panel B of Table 5 reports results from a school-effects specification that uses 2002 data as a control instead of 2000. Because the 2002 treatment and control Bagrut rates are well balanced, these results are larger than those from the 2001–2000 stack. For example, the estimates for girls in the top half of the lagged score distribution climb to about 17 percent, while the estimates for boys in both halves, as well as for girls in the bottom-half sample, are again zero. Finally, estimation using 2000–2002 data produces results between those from the 2000–2001 and 2001–2002 samples, with a modest gain in precision. For example, the estimates for girls in the top half of the lagged score distribution are 0.13–0.14, with a standard error of 0.039.[21]

A further refinement on the models with school effects looks at estimates in the third and fourth quartiles of the lagged score or predicted Bagrut distribution. These results, reported in Table 6, tell a story similar to that in Table 5, except that the largest effects for girls in the stacked models now appear in the upper *quartile* of the classification distribution. For example, when estimated in models without linear control for lagged scores, the effect on girls in the upper (fourth) quartile of the lagged score distribution is 0.145 (s.e. = 0.06), while the corresponding estimate in the third quartile is 0.028 (s.e. = 0.064). This is evidence for the notion that the group most likely to benefit from short-term incentives is those for whom the target is most within reach (although this is apparently a necessary but not sufficient condition, since boys in the upper quartile were unaffected by incentives).

Although the effects on girls in the upper lagged score quartile reported in Table 6 are substantial, these results translate into a modest overall impact after averaging across all treated students (i.e., including both boys and girls and all lagged score quartiles). Table 6 therefore succeeds in using previous test performance to zero in on a subgroup where a fair number of students can be nudged into passing the Bagrut through the exertion of extra effort and a more focused test-taking strategy. This result is consistent with Israeli Ministry of Education reports showing that a substantial number of students—over 20 percent nationally—"almost" get a Bagrut in the sense of fulfilling most but not all Bagrut requirements. For example, many students are only a few

[21] The complete set of logit coefficients corresponding to one of the specifications in panels A and B of Table 5 are reported in Appendix Table A4.

Table 5—Estimates in Covariate Subgroups, Controlling for School Effects

	By lagged score				By predicted probability			
	Boys		Girls		Boys		Girls	
	Top (1)	Bottom (2)	Top (3)	Bottom (4)	Top (5)	Bottom (6)	Top (7)	Bottom (8)
Panel A. 2000 and 2001								
Dependent variable mean	0.341	0.050	0.497	0.074	0.344	0.048	0.498	0.072
Models with:								
School covariates, quartile dummies	−0.043	−0.035	0.093	−0.065	−0.030	−0.069	0.082	−0.046
	[0.045]	[0.039]	[0.043]	[0.035]	[0.045]	[0.042]	[0.043]	[0.035]
School covariates, linear lagged score or predicted prob.	−0.035	−0.031	0.102	−0.052	−0.006	−0.077	0.091	−0.050
	[0.046]	[0.038]	[0.043]	[0.031]	[0.043]	[0.044]	[0.043]	[0.035]
Number of students	2,002	1,395	1,931	1,613	2,001	1,355	1,930	1,639
Panel B. 2001 and 2002								
Dependent variable mean	0.424	0.052	0.569	0.087	0.424	0.051	0.568	0.090
Models with:								
School covariates, quartile dummies	−0.010	0.017	0.165	−0.007	−0.017	0.013	0.144	−0.008
	[0.041]	[0.018]	[0.045]	[0.019]	[0.042]	[0.015]	[0.046]	[0.020]
School covariates, linear lagged score or predicted prob.	−0.001	0.016	0.168	−0.006	−0.013	0.014	0.159	−0.011
	[0.041]	[0.019]	[0.045]	[0.019]	[0.041]	[0.014]	[0.046]	[0.021]
Number of students	2,115	1,532	1,958	1,778	2,115	1,541	1,952	1,782
Panel C. 2000, 2001 and 2002								
Dependent variable mean	0.390	0.047	0.539	0.082	0.390	0.046	0.539	0.080
Models with:								
School covariates, quartile dummies	−0.022	0.009	0.133	−0.019	−0.019	0.000	0.118	−0.015
	[0.038]	[0.018]	[0.039]	[0.020]	[0.039]	[0.017]	[0.039]	[0.020]
School covariates, linear lagged score or predicted prob.	−0.012	0.009	0.139	−0.016	−0.007	0.000	0.129	−0.019
	[0.038]	[0.018]	[0.039]	[0.019]	[0.037]	[0.017]	[0.039]	[0.020]
Number of students	3,137	2,463	2,948	2,692	3,136	2,471	2,941	2,815

Notes: The table reports logit marginal effects estimated in models with school fixed effects. The treatment effect is an interaction between a dummy for treated schools and a dummy for 2001 using data from the years indicated in panel headings. Estimates are for subsamples classified as in Table 4. Robust standard errors are shown in brackets.

units short, while others have the requisite number of units but fail to meet distribution requirements. The next subsection explores the anatomy of Bagrut success in greater detail.[22]

D. *Channels for Improvement*

Students must clear a number of hurdles on the road to Bagrut certification. These include subject tests worth a minimum of 20 or more credit units, a writing composition requirement, and math and English requirements. We therefore looked at the proximate causes of Bagrut success: whether students were tested for more units, were more likely to succeed on the exams they took, and were more likely to satisfy distribution requirements. The first outcome in this context is the number of credit units *attempted*. For example, the basic math curriculum, which awards

[22] At the suggestion of a referee, we estimated models for 2001 that control for lagged (2000) school-average Bagrut rates instead of school fixed effects. As with the estimates reported in Tables 4–5, these results, reported in Appendix Table A5, show no significant effects on boys. The estimates for girls with lagged Bagrut controls are between those generated by models with school effects (Table 5, panel A, repeated in columns 1–4 of Web Table A5) and the larger cross-sectional estimates for 2001 only (Table 4, panel A).

TABLE 6—ESTIMATES BY LAGGED SCORE AND PREDICTED PROBABILITY QUARTILES

	Lagged score quartiles				Predicted probability quartiles			
	Boys		Girls		Boys		Girls	
	4th (1)	3rd (2)	4th (3)	3rd (4)	4th (5)	3rd (6)	4th (7)	3rd (8)
Panel A. 2001								
Dependent variable mean	0.454	0.272	0.616	0.420	0.500	0.227	0.648	0.384
Models with:								
School covariates	−0.025	0.005	0.291	0.143	−0.089	0.005	0.221	0.172
	(0.104)	(0.081)	(0.103)	(0.082)	(0.108)	(0.060)	(0.088)	(0.083)
School covariates, linear	−0.031	0.005	0.299	0.136	−0.091	0.006	0.232	0.177
lagged score or predicted prob.	(0.103)	(0.082)	(0.102)	(0.090)	(0.112)	(0.063)	(0.089)	(0.086)
Number of students	502	478	466	467	508	472	474	458
Panel B. 2000								
Dependent variable mean	0.433	0.203	0.618	0.331	0.427	0.214	0.630	0.325
Models with:								
School covariates	0.078	0.040	0.117	0.088	0.039	0.034	0.089	0.097
	(0.096)	(0.082)	(0.097)	(0.072)	(0.093)	(0.078)	(0.087)	(0.076)
School covariates, linear	0.077	0.040	0.114	0.079	0.004	0.019	0.087	0.099
lagged score or predicted prob.	(0.097)	(0.083)	(0.095)	(0.072)	(0.094)	(0.081)	(0.088)	(0.074)
Number of students	510	512	503	501	511	510	503	499
Panel C. 2000 and 2001 (models with school effects)								
Dependent variable mean	0.443	0.257	0.618	0.370	0.468	0.219	0.641	0.357
Models with:								
School covariates	−0.029	−0.102	0.145	0.028	−0.029	−0.060	0.151	0.041
	[0.061]	[0.078]	[0.060]	[0.064]	[0.065]	[0.062]	[0.063]	[0.073]
School covariates, linear	−0.029	−0.103	0.152	0.038	−0.001	−0.033	0.163	0.047
lagged score or predicted prob.	[0.061]	[0.079]	[0.061]	[0.062]	[0.064]	[0.058]	[0.063]	[0.071]
Number of students	1,010	912	965	922	1,009	981	971	940

Notes: The table reports logit marginal effects in upper-quartile and third-quartile subgroups, classified by lagged test scores or predicted probability of Bagrut success. In panels A and B, BRL standard errors clustered by school are shown in parentheses. Robust standard errors are shown in brackets in panel C.

three units, is completed by taking two tests, one for a single unit, and one for two more units. Students may have responded to program incentives by taking both tests, where they would have previously taken only one.

Estimates of effects on the number of credit units attempted show a small increase for treated girls with lagged scores in the upper half of the 2001 score distribution. This can be seen in Table 7, which reports estimated treatment effects on indicators for units-attempted thresholds for students in the top half of the lagged score distribution. The estimates were constructed using models similar to those used to construct the estimates in Tables 3 and 4, but with different dependent variables. In particular, we look at effects on indicators for at least 18, 20, 22, or 24 units attempted (with attempts measured as of June 2001). The results for girls show a pattern of positive though small and mostly insignificant effects on attempted units. The largest effect, on attempts of 20 or more units, is 0.073 (s.e. = 0.038). Estimated effects on boys' attempts are smaller and none is close to statistical significance.[23]

[23] The estimates for the 2000–2001 levels and 2000–2001 stack in Table 7 use the same sample so they can be more easily compared.

Table 7—Mediating Outcomes

	Boys				Girls			
Outcome variable	2001 Mean (1)	2000 (2)	2001 (3)	2000–2001 (w/school effects) (4)	2001 Mean (5)	2000 (6)	2001 (7)	2000–2001 (w/school effects) (8)
Units attempted								
18	0.749	0.059	0.049	0.012	0.849	0.064	0.091	0.024
		(0.055)	(0.065)	[0.038]		(0.056)	(0.044)	[0.033]
20	0.700	0.050	0.063	0.035	0.793	0.034	0.127	0.073
		(0.063)	(0.066)	[0.040]		(0.054)	(0.053)	[0.038]
22	0.630	0.066	0.052	0.002	0.717	0.042	0.106	0.050
		(0.073)	(0.063)	[0.041]		(0.066)	(0.054)	[0.042]
24	0.536	0.073	0.042	−0.011	0.578	0.033	0.065	0.020
		(0.075)	(0.069)	[0.042]		(0.081)	(0.054)	[0.043]
Units awarded								
18	0.728	0.061	0.057	0.014	0.804	0.097	0.156	0.053
		(0.055)	(0.068)	[0.039]		(0.059)	(0.053)	[0.035]
20	0.686	0.064	0.059	0.016	0.762	0.077	0.150	0.064
		(0.061)	(0.067)	[0.041]		(0.057)	(0.059)	[0.038]
22	0.622	0.052	0.046	0.009	0.688	0.108	0.150	0.049
		(0.072)	(0.062)	[0.043]		(0.065)	(0.058)	[0.041]
24	0.527	0.094	0.046	−0.036	0.590	0.071	0.118	0.045
		(0.080)	(0.065)	[0.044]		(0.079)	(0.066)	[0.043]
Distribution requirements								
Math	0.557	−0.007	0.004	0.020	0.685	0.062	0.153	0.081
		(0.082)	(0.063)	[0.044]		(0.074)	(0.059)	[0.041]
English	0.707	0.107	0.082	−0.001	0.771	0.143	0.111	−0.009
		(0.062)	(0.057)	[0.040]		(0.071)	(0.048)	[0.035]
Writing	0.700	0.014	−0.003	0.002	0.815	−0.002	0.117	0.106
		(0.058)	(0.062)	[0.040]		(0.055)	(0.050)	[0.039]
Bagrut, conditional on units attempted								
18	0.488	0.047	−0.044	−0.085	0.605	0.095	0.200	0.098
		(0.090)	(0.087)	[0.053]		(0.086)	(0.093)	[0.047]
20	0.519	0.050	−0.064	−0.121	0.641	0.128	0.191	0.063
		(0.095)	(0.088)	[0.055]		(0.092)	(0.100)	[0.046]
22	0.556	0.032	−0.055	−0.083	0.664	0.127	0.199	0.054
		(0.106)	(0.100)	[0.058]		(0.097)	(0.104)	[0.046]
24	0.589	0.008	−0.055	−0.065	0.711	0.051	0.176	0.086
		(0.117)	(0.102)	[0.061]		(0.096)	(0.100)	[0.052]

Notes: The table reports logit marginal effects estimated in models with school covariates and a lagged score-group quartile dummy, using data for students in the upper half of the lagged score distribution. Marginal effects for models with school effects were computed using average fitted values for treated students in a sample pooling 2000 and 2001. BRL standard errors are reported in parentheses (for levels estimates). Robust standard errors are reported in brackets (for models with school effects).

Paralleling the increase in the units girls attempted around the 20-unit margin is a small increase in the number of units awarded to girls, as can be seen in the second panel in Table 7.[24] For example, the estimated effect on the probability of obtaining 18-plus units is 0.053 (s.e. = 0.035), while the effect on the likelihood of obtaining 20-plus units is 0.064 (s.e. = 0.038).

[24] These are estimates of effects on indicators for having been awarded 18, 20, 22, and 24 units measured as of June 2002. The timing here has to do with data quality issues. For details, see Angrist and Lavy (2004).

Although these results suggest a modest increase in units awarded, they are too small to explain the program treatment effect on Bagrut rates. It seems likely, therefore, that the net program effect comes partly through other channels.

The means in Table 7 show that 69 percent of high-achieving girls in 2001 obtained 22 credit units. Many therefore fail to get certified because the units are in the wrong subjects. Consistent with this, part of the increase in girls' Bagrut rates appears to have arisen through an increased likelihood of satisfying distribution requirements. These effects are documented in the third panel in Table 7. The estimates with school effects suggest the Achievement Awards program led to a 0.081 (s.e. = 0.041) increase in the likelihood that girls met the math requirement and a 0.11 (s.e. = 0.039) increase in the likelihood that girls met the writing requirement.

The stacked estimates of effects on girls' distribution requirements are stronger than the estimated effects on units attempted and awarded, and might be more indicative of an increase in student effort. By contrast, an increase in units awarded could arise as a mechanical consequence of an increase in units attempted in relatively easy subjects, with little additional student effort. The impact on math and writing requirements seems more likely to be the result of a shift in effort toward these specific subject areas since these are relatively difficult exams. A deeper program impact is also suggested by the increase in the likelihood of Bagrut success conditional on units attempted. The conditional results, reported at the bottom of Table 7, show that treated girls who were tested for at least 18 units were 0.098 (s.e. = 0.047) more likely to obtain a Bagrut, with somewhat smaller effects as the conditioning set gets higher.[25] The effects on distribution requirements and on Bagrut success conditional on attempts suggests that the Achievement Awards program elicited more than a mechanical response involving additional test-taking alone.

Mechanisms and Gender Differences.—Why did girls respond to Achievement Awards while boys did not? We investigated three explanations that seem relevant in our context. The first is that even within classification groups, girls may have been more likely to be "marginal" in the sense of being close to the threshold for a Bagrut. We evaluated this possibility by looking at a number of definitions of near-Bagrut categories in the 2001 control data and in the 2000 data (for example, students with 20 units and all but one of the distribution requirements satisfied). As it turns out, boys and girls were about equally likely to fall into groups that are relatively close to certification.

A second possibility is differences in program awareness. Data from the Ministry of Education follow-up survey, conducted in the late summer and early fall of 2001, can be used to investigate this. The survey data are far from perfect but the results are suggestive. We find that girls are more likely to report having been aware of the program, especially among students in the top half of the lagged score distribution. In this sample, 61 percent of girls and 54 percent of boys demonstrated program awareness. This difference does not seem large enough to explain the boy-girl differences in treatment effects, however. Also, program awareness expressed ex post might simply reflect the higher award rates for girls (since those who got an award are more likely to recall being in the program.)

Finally, and perhaps most relevant, the survey data include measures of study time, overall study effort, and hours worked in paid employment. There is little evidence of a difference in these variables between treatment and control groups. This is true both overall and in analyses by lagged-score subgroups or predicted probability of Bagrut success (a function of parental schooling as well as lagged scores). On the other hand, a common practice in Israeli high schools is for

[25] This estimate is from a model with school effects (column 8). Note that the conditional-on-attempts results do not have a simple causal interpretation when there is also an impact on attempts (see, e.g., Angrist, Eric Bettinger, and Kremer 2006).

students and teachers to get together in marathon study sessions around the holidays (Hanukkah in winter and Passover in spring). The Passover marathon is devoted to a big push for the Bagrut. In our sample, girls are more likely to participate in these marathon study sessions than boys, especially among those in the top half of the lagged score distribution (30 percent for girls versus 19 percent for boys). Among upper-quartile girls, the group with the largest treatment effect on Bagrut rates, we also find significantly higher Passover marathon participation among the treated (an effect of 0.193, s.e. = 0.085). There is no treatment effect on participation in the Hanukkah marathon, a useful specification check since this predates treatment.

The findings on extra study time and distribution requirements suggest that some girls responded to incentives with a focused and well-timed study effort while boys did not. On the other hand, these results do not provide a deep psychological explanation of female responsiveness. It is worth noting, however, that there is a literature suggesting that adolescent girls have more self-discipline (e.g., Angela L. Duckworth and Martin P. Seligman 2006) and are more likely to delay gratification (e.g., Irwin W. Silverman 2003) than adolescent boys. Among young adults, John T. Warner and Saul Pleeter (2001) find that male enlisted personnel behave as if they have higher discount rates than women in the same group.

There is also a consistent pattern of stronger female response to financial incentives in education, with the evidence coming from a surprising variety of settings. Especially relevant are recent studies of tuition aid by Susan Dynarski (2008) and tuition penalties by Pietro Garibaldi et al. (2007), both of which find larger effects for females. Also closely related is a recent randomized trial looking at cash payments for academic achievement among college freshmen; this study finds clear effects for females but no effect on males (Angrist, Daniel Lang, and Oreopoulos 2009). A more modest but still marked gender differential crops up in the response to randomly assigned vouchers for private secondary schools in Colombia (Angrist et al. 2002). These vouchers incorporated an incentive component because voucher retention was conditional on academic performance.[26]

V. Postsecondary Schooling

This section discusses estimates of the effects of the Achievement Awards program on postsecondary school entrollment. The postsecondary academic schooling system in Israel includes seven universities (one of which confers only graduate and PhD degrees), over 40 colleges that confer academic undergraduate degrees (some of these also give master's degrees), and dozens of teachers' and practical engineering colleges that confer bachelor of education or practical engineering degrees.[27] The national enrollment rates for the cohort of graduating seniors in 1995 (through 2001) was 52.7 percent, of which 38 percent were enrolled in universities, 28 percent in general colleges, and 28 percent in teachers' and practical engineering colleges.[28]

The postsecondary outcome variables of interest here are indicators of ever having enrolled in postsecondary institutions of various types as of the 2006–2007 school year. We measure this

[26] Somewhat farther afield, Michael Anderson (2008) shows that three well-known early childhood interventions (Abecedarian, Perry, and the Early Training Project) had substantial short- and long-term effects on girls but no effect on boys. Similarly, a number of public-sector training programs generated larger effects on women than men (Robert J. Lalonde 1995). The Moving to Opportunity randomized evaluation of housing vouchers likewise generated benefits for girls, with little or even adverse effects on boys (Jeffrey R. Kling, Jeffrey B. Leibman, and Lawrence F. Katz 2007).

[27] Practical engineering colleges run two- to three-year programs awarding degrees or certificates in fields like electronics, computers, and industrial production. Two further years of study in an engineering school are required in order to complete a BSc degree in engineering. A 1991 reform sharply increased the supply of postsecondary schooling in Israel by creating publicly funded regional and professional colleges. New institutions granting academic degrees are supervised by the Council for Higher Education, which also supervises the seven research universities.

[28] These data are from the Israel Central Bureau of Statistics, *Report on Post Secondary Schooling of High School Graduates in 1989–1995* (available at: http://www.cbs.gov.il/publications/h_education02/h_education_h.htm).

outcome for our 2000 and 2001 graduating cohorts. Because of compulsory military service, many of the students from these cohorts who enrolled in postsecondary schooling will not have graduated by the 2006–2007 academic year.[29] We therefore focus on enrollment instead of completion.

Our information on postsecondary enrollment comes from administrative records provided by Israel's National Insurance Institute (NII). The NII is responsible for social security and mandatory health insurance in Israel. The NII tracks postsecondary enrollment because students pay a lower health insurance tax rate. Postsecondary schools are therefore required to send a list of enrolled students to the NII every year. For the purposes of our project, the NII Research and Planning Division constructed an extract containing the 2001–2007 enrollment status of students in our study. This file was merged with the other information in our sample and used for analysis at the NII headquarters in Jerusalem.

We coded three indicators for enrollment in postsecondary schooling. The first identifies enrollment in one of the seven universities (at any time from 2001–7); the second expands this definition to include certified academic colleges; the third adds teachers' and practical engineering colleges to the second group. All universities and colleges require a Bagrut for enrollment. Most teachers and practical engineering colleges also require a Bagrut, though some look at specific Bagrut components without requiring full certification.

To avoid an overabundance of results given the many postsecondary outcomes of interest, we focus on results from the specification we review as most reliable. These results, reported in Table 8, are from models similar to equation (3), i.e., estimated using 2000 and 2001 data and including school fixed effects. Estimates with school effects are reported for the top and bottom half of the lagged score distribution as in columns 1–4 of Table 5 (estimates conditional on the probability of Bagrut success are omitted). We also report results from a further split by quartile, as in Table 6. Our discussion focuses on Logit marginal effects, with the exception of effects on university enrollment. Because there are many schools with no students attending university, the results for university enrollment are from linear models only. For comparability, Table 8 reports both OLS and Logit estimates for other outcomes.

Consistent with the fact that few of the students in our study end up in one of Israel's research universities, there is no effect of treatment on university enrollment. This can be seen in panel A of Table 8. University enrollment rates are low even in the top half of the lagged score distribution, 0.064 for boys and 0.069 girls. Enrollment rates in the bottom half of the lagged score distribution are essentially zero for boys and 1.1 percent for girls. The treatment effects for both genders and in both halves of the distribution are also zero. Although university enrollment rates are higher for students in the upper quartile of the lagged score distribution, around 10–11 percent for girls, there is still no program effect on this outcome (as can be seen in columns 5–8).

Combined college and university enrollment rates are much higher than university enrollment rates alone, as shown in panel B of Table 8 (labeled "All Academic"). Panel B also shows some evidence of a program effect on college enrollment for girls in the top quartile of the lagged score distribution. The program effect on college or university enrollment of girls in the 4th quartile is 0.086 (s.e. = 0.055), which can be compared to a mean of 0.248. Moreover, widening the definition of postsecondary study to include teachers' colleges and practical engineering colleges leads to larger and broader effects, reported in panel C. The Logit estimate of the effect on the enrollment of girls using the most inclusive enrollment variable is 0.081 in the top half of the lagged score distribution (s.e = 0.038; mean = 0.331) and 0.128 (s.e. = 0.061) in the top quartile (where the mean is 0.429). These estimates are slightly higher in models with linear control for lagged

[29] Boys serve for three years and girls for two (longer if they take a commission). Ultra-orthodox Jews are exempt from military service as long as they are enrolled in seminary (Yeshiva); orthodox Jewish girls are exempt upon request; Arabs are exempt though some volunteer.

Table 8—Effects on Postsecondary Education by Covariate Subgroup, 2000 and 2001 Data with School Effects

	By lagged score halves				By lagged score quartiles			
	Boys		Girls		Boys		Girls	
	Top (1)	Bottom (2)	Top (3)	Bottom (4)	4th (5)	3rd (6)	4th (7)	3rd (8)
Panel A. University								
Dependent variable mean	0.064	0.003	0.069	0.011	0.101	0.025	0.112	0.027
Estimation method:								
OLS	0.019	0.004	−0.021	0.014	0.041	0.014	0.012	−0.034
	[0.022]	[0.004]	[0.023]	[0.008]	[0.038]	[0.022]	[0.042]	[0.020]
Panel B. All academic								
Dependent variable mean	0.171	0.041	0.176	0.046	0.230	0.110	0.248	0.107
Estimation method:								
OLS	0.009	0.001	0.028	−0.006	0.002	0.035	0.086	−0.031
	[0.033]	[0.018]	[0.033]	[0.020]	[0.051]	[0.040]	[0.055]	[0.039]
Logit	0.022	0.005	0.045	0.018	0.009	0.045	0.086	−0.006
	[0.026]	[0.019]	[0.033]	[0.019]	[0.046]	[0.030]	[0.055]	{0.065}
Panel C. Academic, teachers, and practical engineering colleges								
Dependent variable mean	0.252	0.081	0.331	0.099	0.304	0.198	0.429	0.236
Estimation method:								
OLS	−0.041	−0.018	0.067	0.035	−0.042	−0.042	0.123	0.031
	[0.038]	[0.025]	[0.040]	[0.028]	[0.055]	[0.052]	[0.060]	[0.052]
Logit	−0.028	−0.012	0.081	0.036	−0.032	−0.031	0.128	0.047
	[0.035]	[0.028]	[0.038]	[0.019]	[0.051]	[0.056]	[0.061]	[0.051]
Panel D. Bagrut (replication using the sample with National Insurance data)								
Dependent variable mean	0.341	0.034	0.490	0.070	0.443	0.236	0.611	0.372
Estimation method:								
OLS	−0.027	−0.011	0.102	−0.049	−0.022	−0.063	0.161	0.038
	[0.039]	[0.016]	[0.042]	[0.024]	[0.060]	[0.051]	[0.060]	[0.059]
Logit	−0.038	−0.038	0.108	−0.053	−0.029	−0.116	0.163	0.037
	[0.046]	{0.040}	{0.043}	[0.032]	[0.061]	[0.080]	{0.062}	{0.060}
Number of students	1,997	1,985	1,882	1,711	1,007	988	921	958

Notes: The table reports OLS estimates and Logit marginal effects for models using 2000 and 2001 data with school effects. All models include school covariates and linear lagged score controls. Columns 1–4 show estimates in top and bottom subgroups as in Table 5. Columns 5–8 show estimates in upper and third-quartile subgroups as in Table 6. Robust standard errors are reported in brackets. Conventional standard errors are reported in braces where robust Logit standard error estimation failed. Where both are available, the conventional and robust standard errors are virtually identical. The reported means and number of observations are for the OLS samples.

scores. In contrast, the estimates for boys are mostly small and insignificant for all enrollment outcomes in all subsamples.

Program effects on the most inclusive enrollment outcome in the upper half and upper quartile of the lagged score distribution are about three-fourths the size of the program effects on Bagrut rates when the latter are estimated using the same model and sample. For purposes of this comparison, Bagrut results for the relevant sample appear at the bottom of Table 8. These estimates are slightly smaller than those reported in Table 5 (for top half girls) and slightly larger than those reported in Table 6 for upper quartile girls).

The Bagrut results in Table 8 suggest that three-quarters of the additional Bagrut rates received as a consequence of the Achievement Awards intervention caused additional postsecondary enrollment of some type (e.g., compare 0.12 in panel C of column 7 with 0.16 in panel D). Although this proportion may seem high, it bears emphasizing that the overall postsecondary enrollment rate for all Israeli Bagrut holders is also high, on the order of 78 percent for the 1994/95 cohort.

A related point is the fact that noncompliance (the fact that some treated schools did not participate) does not affect the ratio of postsecondary effects to Bagrut effects. This ratio can be interpreted as an instrumental variables estimate of the effect of Bagrut certification on postsecondary enrollment. The instrumental variables adjustment for noncompliance implicitly divides both postsecondary and Bagrut results by the same take-up rate. On the other hand, the program impact on certification is not the only channel by which the Achievement Awards program may have increased postsecondary enrollment. In other words, Bagrut certification need not satisfy an exclusion restriction for the reduced-form program effect on postsecondary outcomes. Some students may have had better postsecondary options by virtue of increasing the number of units tested or by satisfying distribution requirements, even if their Bagrut status was unaffected.

VI. Summary and Conclusions

A randomly assigned offer of cash awards to students in low-achieving schools appears to have generated substantial increases in the matriculation rates of girls. Although there is some imbalance in Bagrut rates from the year preceding treatment, a causal interpretation of the results is supported by estimates from models that control for unobserved school effects, and by the absence of a treatment effect in the cohort that graduated after the one treated. The overall impact on girls is driven by treatment effects in a group we see as marginal; that is, students relatively close to certification thresholds. The effect on this group (girls in the upper half of the lagged score distribution) are around 0.10 in a model allowing for omitted school effects. This is our best guess of the program impact on the marginal group of girls. There appears to be no effect on boys in general or on girls who are not in the marginal group, so that the overall program effect was small. It also seems worth mentioning that we found no evidence of negative or perverse effects.

The estimated impact on girls increases when allowance is made for the fact that one-quarter of schools offered the opportunity to participate in the program either declined to participate or failed to provide rosters in time. Adjusting for noncompliance, the effect on treated girls (i.e., girls in treated schools) is about one-third larger (1/0.75) than the reduced-form intention-to-treat effects discussed in the paper. On the other hand, the intention-to-treat effect may be a better gauge of future program impact if other programs of this sort give school administrators the opportunity to opt out.

An analysis of the channels through which students may have responded to incentives generates some evidence of increased effort in the form of more exams taken or more difficult exams attempted and, especially, an increased likelihood of meeting distribution requirements. This turns up in a higher success rate for girls conditional on the number of exams attempted. Using survey data, we also find girls increased their exam study time in the pre-Bagrut holiday period. Boys did not respond in any way that we can detect. The gender differential in program response echoes male-female differences in the response to financial incentives for college achievement and in the response to tuition subsidies and penalties in a number of recent studies.

We have also shown that, for many students, the increase in matriculation rates translated into increased postsecondary enrollment. The sharpest boost is for girls in the top quartile of the lagged score distribution, a 12–13 percentage point increase, while the postsecondary enrollment gain for girls in the top half of the lagged score distribution is about 7 percentage points. These increases seem likely to have generated substantial economic gains since the returns to postsecondary education in Israel appear to be high. For example, R. Frisch and J. Moalem (1999) estimate the return to a year of college to be about 11 percent in the late 1990s, while Frisch (2007) estimates the average return to having any postsecondary schooling to be about 34 percent.[30]

Because of the substantial economic return to postsecondary education, the Achievement Awards incentive scheme is likely to have generated a net social gain. The bonus offer of NIS6,000 shekels for matriculating seniors was worth about $1,429 at the time the treated cohort finished school. About 27 percent of the treatment group received bonuses, so the cost was roughly $385 per treated student. Using the average annual earnings of those with 11–12 years of schooling in 2005 as a base, ($14,910), and assuming that it takes 10 years before any benefits are realized (to allow for military service, college attendance, and labor-market entry), the internal rate of return for investment in Achievement Awards is about 7.3 percent.[31] This suggests that cash incentives of this sort can make economic sense even without taking distributional implications into account. More focused programs, e.g., programs targeting girls and/or marginal students, could well generate even higher economic returns.

Finally, it bears emphasizing that the Achievement Awards demonstration focused on cash incentives for achievement on a high-stakes exam. A natural question for future research is how the results of this sort of intervention compare with those from closely related policies, such as interventions that pay students for school attendance *per se* instead of or in addition to achievement (e.g., conditional cash transfers as in Mexico's Progresa and the Education Maintenance Allowance in Britain). Since other interventions have so far involved different populations and targeted different endpoints, a careful comparison will most likely require new evaluations and a long horizon for data collection. It seems likely that the most informative cross-program comparisons will be based on long-run outcomes such as postsecondary enrollment, college completion, and ultimately, earnings.

Appendix

Program and Data Collection Time Line

Schools selected, treatment assigned, and treated principals contacted	December	2000
Orientation for principals and students	January	2001
Baseline administrative data collected	January	
Spring study marathons	March–April	
Media coverage	May	
Bagrut tests	June	
Student survey	August–October	
Retest (math and English)	August–September	
Winter retest	December–January	2001–2002

Notes: In March 2001 principals were interviewed to determine whether the program was publicized in schools. Bonuses were paid in May 2002.

[30] Of course, the economic returns to schooling for affected students in our study may be higher or lower than 11 percent. For example, Card (1995) argues that the returns to schooling for credit-constrained students should be higher than population average returns.

[31] This calculation is based on the following assumptions: The estimates in Table 8 suggest the program raised college attendance by say 0.075 in the top half of the girls lagged score distribution (taking a number between logit and OLS for the broadest category). This implies an average enrollment effect (including boys) of 0.01875. Assuming the schooling generated by this enrollment boost amounts to two years, each yielding an 11 percent rate of return, the program effect is worth $0.01875 \times 0.11 \times 2 = 0.004125$ percent. Based on average annual earnings of $14,910, the gain per year is $61.5. The earnings data are from http://www.cbs.gov.il/publications/income_05/pdf/t22.pdf.

Table A1—School Average Bagrut Rates

Pair	Treated	Non-complier	Arab school	Relig. school	Enrollment 1999	Enrollment 2000	Enrollment 2001	Enrollment 2002	Bagrut passing rate 1999	Bagrut passing rate 2000	Bagrut passing rate 2001	Bagrut passing rate 2002
1			X		153	173	175	249	0.046	0.000	0.091	0.185
1	X			X	56	59	45	43	0.036	0.051	0.000	0.047
2				X	242	170	147	88	0.054	0.094	0.184	0.034
2	X				179	185	145	158	0.050	0.108	0.110	0.095
3					88	99	71	80	0.114	0.000	0.056	0.075
3	X		X		123	129	99	103	0.098	0.054	0.030	0.068
4					81	68	73	67	0.148	0.162	0.082	0.075
4	X		X		187	223	248	297	0.134	0.390	0.339	0.458
5					125	124	96	70	0.152	0.105	0.083	0.129
5	X			X	55	39	38	48	0.145	0.077	0.579	0.167
6	X				117	125	123	154	0.171	0.136	0.154	0.273
7				X	16	28	16	22	0.188	0.214	0.375	0.545
7	X			X	67	85	58	63	0.179	0.165	0.483	0.444
8				X	57	48	61	60	0.193	0.771	0.328	0.583
8	X				90	97	113	106	0.189	0.186	0.168	0.368
9					61	40	59	60	0.197	0.350	0.000	0.383
9	X			X	10	14	9	21	0.200	0.071	0.667	0.429
10				X	34	39	26	43	0.206	0.410	0.654	0.488
10	X	X			135	135	108	102	0.207	0.267	0.361	0.441
11					136	148	134	169	0.213	0.176	0.164	0.172
11	X				129	158	152	159	0.209	0.165	0.092	0.151
12			X		19	24	20	60	0.211	0.667	0.250	0.617
12	X			X	32	44	24	20	0.219	0.250	0.500	0.350
13					146	118	119	137	0.219	0.153	0.185	0.219
13	X				85	80	86	114	0.224	0.363	0.372	0.342
14					208	170	185	199	0.236	0.153	0.276	0.352
14	X	X	X		75	50	64	120	0.227	0.560	0.484	0.367
15			X		156	153	163	202	0.244	0.176	0.331	0.391
15	X	X			138	141	152	171	0.254	0.610	0.467	0.520
16			X		102	115	108	107	0.255	0.226	0.213	0.327
16	X				74	61	75	74	0.257	0.098	0.107	0.095
17				X	23	14	16	0	0.261	0.071	0.000	—
17	X		X		76	68	67	62	0.263	0.441	0.448	0.435
18			X		216	211	219	246	0.273	0.303	0.301	0.305
18	X	X			200	148	110	146	0.275	0.162	0.173	0.103
19					141	111	77	183	0.284	0.541	0.636	0.776
19	X	X			123	40	62	43	0.276	0.025	0.081	0.093
20					185	161	111	94	0.286	0.161	0.126	0.181
20	X		X		144	144	167	188	0.285	0.389	0.353	0.309

Notes: The table reports statistics for each school in the Achievement Awards demonstration. The control school in pair 6 closed before treatment assignments were announced (in models with pair effects, the treated school in pair 6 is assigned to pair 7). Noncompliant schools are treated schools that did not participate in the program. Pairs were constructed using 1999 data.

Table A2—Covariate Balance

	All		Boys		Girls	
	Mean (1)	Difference (2)	Mean (3)	Difference (4)	Mean (5)	Difference (6)
Panel A. 2001						
School covariates						
Arab school	0.348	−0.034	0.374	−0.147	0.320	0.071
		[0.191]		[0.202]		[0.198]
Religious school	0.115	−0.052	0.084	0.093	0.148	−0.190
		[0.096]		[0.076]		[0.138]
Micro covariates						
Father's education	10.1	0.365	9.82	1.31	10.3	−0.490
		[0.698]		[0.875]		[0.631]
Mother's education	10.0	0.587	9.87	1.45	10.2	−0.219
		[0.872]		[1.03]		[0.839]
Number of siblings	3.74	0.097	3.65	−0.110	3.84	0.362
		[0.733]		[0.748]		[0.784]
Immigrant	0.064	−0.059	0.029	0.019	0.100	−0.126
		[0.072]		[0.015]		[0.120]
Lagged score	53.1	1.17	52.1	−0.223	54.2	3.18
		[4.51]		[4.86]		[6.32]
Observations	3,821		1,960		1,861	
Panel B. 2000						
Bagrut rate	0.224	0.048	0.177	0.041	0.272	0.083
		[0.055]		[0.053]		[0.072]
School covariates						
Arab school	0.319	−0.032	0.352	−0.131	0.286	0.050
		[0.181]		[0.196]		[0.184]
Religious school	0.134	−0.029	0.098	0.096	0.170	−0.139
		[0.106]		[0.092]		[0.149]
Micro covariates						
Father's education	9.9	0.328	9.75	1.27	10.0	−0.557
		[0.716]		[0.922]		[0.695]
Mother's education	9.8	0.536	9.71	1.58	9.9	−0.459
		[0.882]		[1.06]		[0.867]
Number of siblings	3.68	0.150	3.53	0.015	3.84	0.372
		[0.629]		[0.621]		[0.676]
Immigrant	0.074	−0.053	0.039	0.012	0.109	−0.102
		[0.067]		[0.027]		[0.105]
Lagged score	50.2	4.48	49.1	5.052	51.4	4.63
		[4.71]		[4.73]		[7.15]
Observations	4,039		2,038		2,001	

Notes: This table reports means and treatment-control differences by gender in 2001 (the treatment year) and 2000 (the pre-treatment year). Standard errors, clustered by school, are reported in brackets.

Table A3—Comparison of Grouped and Micro Data Estimates

	Two-step procedure (grouped estimates)						Micro data estimates		
	Unweighted			Weighted					
Pair effects	Boys + girls (1)	Boys (2)	Girls (3)	Boys + girls (4)	Boys (5)	Girls (6)	Boys + girls (7)	Boys (8)	Girls (9)
Panel A. 2001									
No	0.106	0.036	0.114	0.055	−0.003	0.095	0.055	−0.004	0.095
	(0.048)	(0.050)	(0.048)	(0.038)	(0.042)	(0.045)	(0.036)	(0.042)	(0.043)
	[0.046]	[0.047]	[0.047]	[0.036]	[0.042]	[0.044]	{0.039}	{0.046}	{0.047}
Yes	0.111	—	—	0.064	—	—	0.064	—	—
	(0.051)			(0.040)			(0.025)		
	[0.034]			[0.025]			{0.036}		
Panel B. 2000									
No	−0.007	0.000	0.047	0.032	0.031	0.059	0.031	0.031	0.058
	(0.049)	(0.052)	(0.056)	(0.043)	(0.046)	(0.051)	(0.041)	(0.048)	(0.044)
	[0.047]	[0.054]	[0.052]	[0.041]	[0.048]	[0.045]	{0.043}	{0.052}	{0.048}
Yes	−0.003	—	—	0.040	—	—	0.040	—	—
	(0.051)			(0.046)			(0.030)		
	[0.033]			[0.031]			{0.045}		

Notes: All models control for school covariates and lagged score quartile dummies. Columns 1–6 report estimates constructed using the Donald and Lang (2007) two-step procedure, where the first step estimates school effects (means) adjusted for micro covariates, and the second step is a group-level regression using the adjusted means. The micro covariates in the first step are dummies for lagged score quartiles. Conventional standard errors for the second step are shown in parentheses. Heteroskedasticity-consistent standard errors for the second step are reported in brackets. Columns 7–9 report regression results using micro data. The standard errors in parentheses in columns 7–9 are adjusted for school clustering using formulas in Liang and Zeger (1986). Standard errors in braces in columns 7–9 use Bell and MacCaffrey's (2002) BRL estimator.

Table A4—Parameter Estimates for Models in Table 5
(*Linear control, top half sample, by lagged score*)

	2000 and 2001		2001 and 2002	
	Boys	Girls	Boys	Girls
Dependent variable mean	0.341	0.497	0.424	0.569
Treated	−0.173	0.527	−0.006	0.840
	[0.226]	[0.220]	[0.220]	[0.225]
Arab school × (year = 2001)	−0.175	0.254	−0.198	0.377
	[0.227]	[0.236]	[0.222]	[0.230]
Religious school × (year = 2001)	1.76	−0.354	1.22	1.17
	[0.376]	[0.355]	[0.365]	[0.371]
Lagged score	0.048	0.073	0.062	0.062
	[0.006]	[0.007]	[0.006]	[0.007]
(Year = 2001)	0.118	−0.409	−0.519	−0.958
	[0.206]	[0.195]	[0.200]	[0.195]
Number of students	2,002	1,931	2,115	1,958

Notes: The table reports logit coefficients for models in columns 1–4 of Table 5 (panels A and B). Robust standard errors are shown in brackets.

Table A5—Estimates with Lagged Bagrut Rates Compared to Estimates with School Effects

	2000 and 2001 with school effects				2001 including lagged Bagrut			
	Boys		Girls		Boys		Girls	
	Top (1)	Bottom (2)	Top (3)	Bottom (4)	Top (5)	Bottom (6)	Top (7)	Bottom (8)
Dependent variable mean	0.341	0.050	0.497	0.074	0.365	0.035	0.518	0.056
Models with:								
School covariates, lagged score quartile dummies	−0.043	−0.035	0.093	−0.065	−0.021	−0.005	0.153	−0.035
	[0.045]	[0.039]	[0.043]	[0.035]	(0.069)	(0.020)	(0.077)	(0.026)
School covariates, linear lagged score	−0.035	−0.031	0.102	−0.052	−0.019	−0.005	0.162	−0.031
	[0.046]	[0.038]	[0.043]	[0.031]	(0.068)	(0.020)	(0.079)	(0.026)
Number of students	2,002	1,395	1,931	1,613	980	980	933	928

Notes: The table reports logit marginal effects. Columns 1–4 repeats the 2000/2001 results from Table 5, with robust standard errors in brackets. Columns 5–8 show estimates from 2001, including a control for the lagged Bagrut mean (in 2000), estimated by school, gender, and lagged-score quartile, with BRL standard errors in parentheses. Top and bottom subgroups were constructed using lagged scores.

REFERENCES

Anderson, Michael. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103(484): 1481–95.

Angrist, Joshua, Eric Bettinger, Erik Bloom, Beth King, and Michael Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *American Economic Review*, 92(5): 1535–58.

Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2006. "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *American Economic Review*, 96(3): 847–62.

Angrist, Joshua, and Jinyong Hahn. 2004. "When to Control for Covariates? Panel Asymptotics for Estimates of Treatment Effects." *Review of Economics and Statistics*, 86(1): 58–72.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2009. "Incentives and Services for College Achievement: Evidence from a Randomized Trial." *American Economic Journal: Applied Economics*, 1(1): 136–63.

Angrist, Joshua D., and Victor Lavy. 2004. "The Effect of High Stakes High School Achievement Awards: Evidence from a School-Centered Randomized Trial." IZA Working Paper 1146.

Behrman, Jere R., Piyali Sengupta, and Petra Todd. 2000. *Final Report: The Impact of PROGRESA on Achievement Test Scores in the First year.* Washington, DC: International Food Policy Research Institute.

Bell, Robert M., and Daniel F. McCaffrey. 2002. "Bias Reduction in Standard Errors for Linear Regression with Multi-stage Samples." *Survey Methodology*, 28(2): 169–181.

Bloom, Dan, and Colleen Sommo. 2005. "Building Learning Communities: Early Results from the Opening Doors Demonstration at Kingsborough Community College." New York: MDRC. http://www.mdrc.org/publications/410/overview.html.

Card, David E. 1995. "Earnings, Schooling, and Ability Revisited." *Research in Labor Economics,* 14: 23–48.

Central Bureau of Statistics. 2002. *Statistical Abstract of Israel 53*. Jerusalem: Central Bureau of Statistics.

Dee, Thomas S., and Brian A. Jacob. 2007. "Do High School Exit Exams Influence Educational Attainment or Labor Market Performance?" In *Standards-Based Reform and Children in Poverty: Lessons for "No Child Left Behind,"* ed. Adam Gamoran, 154–200. Washington, DC: Brookings Institution Press.

Diehr, Paula H., Donald C. Martin, Thomas Koepsell, and Allen Cheadle. 1995. "Breaking the Matches in a Paired t-Test for Community Interventions When the Number of Pairs is Small." *Statistics in Medicine,* 14(13): 1491–1504.

Donald, Stephen G., and Kevin Lang. 2007. "Inference with Difference-in-Differences and Other Panel Data." *Review of Economics and Statistics*, 89(2): 221–33.

Donner, Allan, K., Stephen Brown, and Penny Brasher. 1990. "A Methodological Review of Non–Therapeutic Intervention Trials Employing Cluster Randomization, 1979–1989." *International Journal of Epidemiology,* 19: 795–800.

Duckworth, Angela Lee, and Martin P. Seligman. 2006. "Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores." *Journal of Educational Psychology,* 98(1): 198–208.

Dynarski, Mark, and Gleason, Philip. 2002. "How Can We Help? What We Have Learned From Recent Federal Dropout Prevention Evaluations." *Journal of Education for Students Placed at Risk*, 7(1): 43–69.

Dynarski, Susan. 2008. "Building the Stock of College-Educated Labor." *Journal of Human Resources*, 43(3): 576–610.

Eckstein, Zvi, and Kenneth I. Wolpin. 1999. "Why Youths Drop out of High School: The Impact of Preferences, Opportunities, and Abilities." *Econometrica*, 67(6): 1295–339.

Feng, Ziding, Paula H. Diehr, A. Peterson, and Dale McLerran. 2001. "Selected Statistical Issues in Group Randomized Trials." *Annual Review of Public Health,* 22: 167–87.

Frisch, R. 2007. "The Return to Schooling -the Causal Link Between Schooling and Earnings." Bank of Israel Working Paper 2007.03.

Frisch, R., and J. Moalem. 1999. "The Rise in the Return to Schooling in Israel in 1976–1997." Bank of Israel Working Paper 99.06.

Fuller, Winship C., Charles F. Manski, and David A. Wise. 1982. "New Evidence on the Economic Determinants of Postsecondary Schooling Choices." *Journal of Human Resources*, 17(4): 477–98.

Gail, Mitchell H., Steven D. Mark, Raymond J. Carroll, Sylvan B. Green, and David Pee. 1996. "On Design Considerations and Randomization-based Inference for Community Intervention Trials." *Statistics in Medicine,* 15: 1069–92.

Garibaldi, Pietro, Francesco Giavazzi, Andrea Ichino, and Enrico Rettore. 2007. "College Cost and Time to Complete a Degree: Evidence from Tuition Discontinuities" National Bureau of Economic Research Working Paper 12863.

Gruber, Jonathan, ed. 2001. *Risky Behavior among Youths: An Economic Analysis*. Chicago: University of Chicago Press.

Heckman, James, and Edward Vytlacil. 1998. "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return Is Correlated with Schooling." *Journal of Human Resources*, 33(4): 974–87.

Hotz, V. Joseph, Lixin Colin Xu, Marta Tienda, and Avner Ahituv. 2002. "Are There Returns to the Wages of Young Men from Working While in School?" *Review of Economics and Statistics*, 84(2): 221–36.

Israel Ministry of Education. 2001. *Statistics of the Matriculation Examination* (*Bagrut*) *Test Data, 2000*. Jerusalem: Ministry of Education Chief Scientist's Office.

Israel Ministry of Education. 2002. *The Bagrut 2001 Program: An Evaluation*. Jerusalem: Ministry of Education Evaluation Division.

Jackson, C. K. 2007. "A Little Now for a Lot Later: A Look at a Texas Advanced Placement Incentive Program." http://works.bepress.com/c_kirabo_jackson/1.

Kane, Thomas J., and Douglas O. Staiger. 2002. "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives*, 16(4): 91–114.

Keane, Michael P., and Kenneth I. Wolpin. 2000. "Eliminating Race Differences in School Attainment and Labor Market Success." *Journal of Labor Economics*, 18(4): 614–52.

Kohn, Alfie. 1999. *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes*. Boston: Houghton Mifflin.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75(1): 83–119.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. Forthcoming. "Incentives to Learn." *The Review of Economics and Statistics.*

LaLonde, Robert J. 1995. "The Promise of Public Sector-Sponsored Training Programs." *Journal of Economic Perspectives*, 9(2): 149–68.

Lang, Kevin. 1993. "Ability Bias, Discount Rate Bias and the Return to Education." http://people.bu.edu/lang/ability.pdf.

Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw. 2003. "The Effect of Financial Rewards on Students' Achievements: Evidence from a Randomized Experiment." Center for Economic Policy Research Discussion Paper 3921.

Leuven, Edwin, Hessel Oosterbeek, J. Sonnemans, and Bas van der Klaauw. 2008. “Incentives Versus Sorting in Tournaments: Evidence from a Field Experiment.” IZA Discussion Paper 3326.

Liang, Kung-yee, and Scott L. Zeger.1986. “Longitudinal Data Analysis Using Generalized Linear Models.” Biometrika, 73(1): 13–22.

Long, David M., J.M. Gueron, R.G.Wood, R. Fisher, and V. Fellerath. 1996. *LEAP: Three–year Impacts of Ohio's Welfare Initiative to Improve School Attendance Among Teenage Parents*. New York: MDRC.

MacKinnon, James G., and Halbert White. 1985. “Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties.” *Journal of Econometrics*, 29(3): 305–25.

Martorell, Francisco. 2005. “Do High School Graduation Exams Matter? Evaluating the Effects of Exit Exam Performance on Student Outcomes.” Unpublished.

Maxfield, Myles, Allen Schirm, and Nuria Rodriguez-Planas. 2003. “The Quantum Opportunities Program Demonstration: Implementation and Short-Term Impacts.” Mathematica Policy Research Report 8279–093.

Moulton, Brent R. 1986. “Random Group Effects and the Precision of Regression Estimates.” *Journal of Econometrics*, 32(3): 385–97.

Reich, Robert. 1998. “Putting the Surplus, if Any, to Work.” *The New York Times,* January 9. http://query.nytimes.com/gst/fullpage.html?res=9B06E4DF1F30F93AA35752C0A96E958260&scp=5&sq=robert%20B.%20reich%201998&st=cse.

Schultz, T. Paul. 2004. “School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program.” *Journal of Development Economics*, 74(1): 199–250.

Silverman, Irwin W. 2003. “Gender Differences in the Delay of Gratification: A Meta-Analysis.” *Sex Roles,* 49(9–10): 451–63.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2003. “Working During School and Academic Performance.” *Journal of Labor Economics*, 21(2): 473–91.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2004. “Time-Use and College Outcomes.” *Journal of Econometrics*, 121(1–2): 243–69.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2008. “The Causal Effect of Studying on Academic Performance.” *B.E. Journal of Economic Analysis and Policy: Frontiers of Economic Analysis and Policy*, 8(1): Article 14.

Thornquist, Mark D., and Garnet L. Anderson. 1992. *Small-Sample Properties of Generalized Estimating Equations in Group-Randomized Designs with Gaussian Response.* Seattle: Fred Hutchinson Cancer Research Center Technical Report.

Twersky, Isadore. 1972. *A Maimonides Reader.* Springfield, NJ: Behrman House, Inc.

Tyler, John H. 2003. “Using State Child Labor Laws to Identify the Effect of School-Year Work on High School Achievement.” *Journal of Labor Economics*, 21(2): 381–408.

Warner, John T., and Saul Pleeter. 2001. “The Personal Discount Rate: Evidence from Military Downsizing Programs.” *American Economic Review*, 91(1): 33–53.

Wooldridge, Jeffrey M. 2003. “Cluster-Sample Methods in Applied Econometrics.” *American Economic Review*, 93(2): 133–38.